GUEST EDITORIAL

Revising Manuscripts and Grants.
Karen Hassey Dow, PhD, RN, FAAN

LETTER TO THE EDITOR

Response to Mary Blegen’s Editorial in the March/April Issue of Nursing Research.
Melanie C. Dreher

FEATURE

Efficacy of Theory-Based Activities for Behavioral Symptoms of Dementia.
Ann M. Kolanowski; Mark Litaker; Linda Buettner

An Analytic Strategy for Modeling Multiple-Item Responses: A Breast Cancer Symptom Example.
Ardith Z. Doorenbos; Natalya Verbitsky; Barbara Given; Charles W. Given

Dating Violence in College Women: Associated Physical Injury, Healthcare Usage, and Mental Health Symptoms.
Angela Frederick Amar; Susan Gennaro

The Effectiveness of a Nursing Inpatient Smoking Cessation Program in Individuals With Cardiovascular Disease.
Maud-Christine Chouinard; Sylvie Robichaud-Ekstrand

A Meta-Analysis of the Effect of Hospital-Based Case Management on Hospital Length-of-Stay and Readmission.
Young-Ju Kim; Karen L. Soeken

Erratum

A Randomized Clinical Trial of an HIV-Risk-Reduction Intervention Among Low-Income Latino Women.

FEATURE

Testing the Psychometric Properties of the Minnesota Living With Heart Failure Questionnaire.
Seongkum Heo; Debra K. Moser; Barbara Riegel; Lynne A. Hall; Norma Christman

Predicting Nosocomial Bloodstream Infections Using Surrogate Markers of Injury Severity: Clinical and Methodological Perspectives.
Maher M. El-Masri; Tarek A. Hammad; Susan M. Fox-Wasylyshyn

METHOD

Use of Time Studies for Determining Intervention Costs.
Mary J. Findorff; Jean F. Wyman; Catherine F. Croghan; John A. Nyman
As a researcher and author, how do you feel and what do you do when the editor requests revisions on a manuscript you submitted for publication? Recently my coauthors and I received reviews and editorial feedback with a number of constructive recommendations. We outlined the reviewers’ comments and revised the manuscript, carefully addressing each point. This unremarkable reaction to requests for revisions is a normal activity for authors. Whether the manuscript is accepted or not, the reviewers’ comments and editorial remarks are carefully and thoughtfully considered. Critique is taken in a spirit of collegiality to improve the final product. In short, we tend to accept and use the feedback we get.

Now, consider how we feel and respond when we have a grant application that receives a poor score or is not scored. I thought about this recently when a colleague—a well-funded researcher and grant reviewer—received notification of an “unscored” application signifying that the grant ranked in the lower half. The reaction was anger directed at study section reviewers. I have seen similar reactions in other researchers, both novice and seasoned. To my colleague the reviewers were, at best, not competent to review the proposal; at worst, they were short-sighted, blind, ignorant, and feeble. Hardly the stuff of respectful collegial exchange.

Anger and disappointment in receiving an unscored application are understandable. Countless hours of hard work and effort are expended in the submission process. Unfortunately, when anger and resentment creep into a revised application, reviewers may interpret such content as unresponsive to the reviews, adversarial, and, in many cases, arrogant. Rarely does anger directed toward the reviewers contribute to a better score.

Study section members evaluate and critique an application on the basis of (a) scientific rigor, (b) significance, (c) innovation, (d) design and methods, and (e) adequacy of the research team and research environment. They determine the relative strengths and weaknesses of applications in considering a score. They do not make funding decisions. They do, however, contribute countless hours in grant reviews, hours they could have spent writing a manuscript or preparing their own grant application. And, while some reviewers may have weak eyesight (perhaps from reading applications containing less than 11-point font size), none, to my knowledge, are short-sighted or blind.

Does the researcher’s attitude toward a critique make a difference? Perhaps, the next time you receive a not-so-favorable score and revise an application, consider your grant reviewers as 2 or 3 real-life colleagues sitting before you. Think of them as having your best interest at heart, guiding you to strengthen your application, and contributing to your scholarly development. Place anger outside of the revision. Alternatively, think about the last time that you revised and resubmitted a manuscript, carefully incorporating the sage and wise comments of reviewers. Recall the satisfaction of receiving favorable reviews that culminated in a publication.

Karen Hassey Dow, PhD, RN, FAAN
Associate Editor
Mary Blegen has eloquently and convincingly identified the extraordinary benefits of three decades of nursing research, guided and proliferated by the National Institute for Nursing Research (NINR). She is so right. As a member of the charter study section of what was then the National Center for Nursing Research, I have been privileged to witness our expanded nursing knowledge buttress and hone nursing practice. Doctoral programs in nursing have proliferated, research journals have matured, and despite concerns about the future of the nursing professoriate, we have faculties with credentials that are consistent with the tradition of the academy.

The potent role of evidence in practice is plain, but evidence, alone, does not yield better clinical outcomes. The translation of evidence into practice requires highly qualified and well-credentialed clinicians, in leadership roles, to introduce and institutionalize practice innovations. An irony of the current debate is that the Doctor of Nursing Practice (DNP) is truly the logical and inevitable product of our researchers, who have expanded our knowledge base so bountifully that it cannot be encompassed in the conventional time and credit allotment for master’s preparation. The tenor of the DNP controversy also suggests an intraprofessional dissonance between the practitioners and researchers, despite the experience of other professions in which the advancement of the discipline is related to having both qualified practitioners and scientists, working in an interdependent and mutually respectful way.

Many of the arguments in opposition to the DNP assume a zero sum game in which the practice doctorate will usurp the limited pool of research doctorate aspirants, but where is the evidence? This is not unlike the argument, 25 years ago, that NINR would appropriate resources desperately needed by the Division of Nursing. Indeed, another irony of the debate is that the protestors’ commitment to evidence for guiding practice has not been applied to education. A competing assumption to the “limited good” concept might be that the leadership opportunities that accompany the title of Doctor, enjoyed by practitioners in all health professions except nursing, will attract and retain quality, science-oriented individuals who would otherwise have chosen medicine, dentistry, pharmacy, physical therapy, etc., and further, that many of these may pursue a research doctorate.

As the repository of brilliant women with limited career choices, nursing has been able, with persistence and intelligence, to meet professional and intellectual challenges, despite enormous odds. But we can no longer rely on sex discrimination to assure the intellectual capital needed for nursing excellence. To attract the best and brightest to be our future scientists we must engage them. Rather than trying to preserve the scientific integrity of the profession by limiting options for clinicians, we must resolve to creatively expand our pool of potential scientists with direct-entry PhD programs, undergraduate research opportunities, funding students’ participation in research meetings, coauthoring papers, mentoring, etc. If the PhD is so fragile that it is threatened by the long overdue appropriate credentialing of qualified practitioners, we have more to worry about in nursing than a new title.

Melanie C. Dreher
Kelting Dean and Professor
The University of Iowa College of Nursing
E-mail: melanie-dreher@uiowa.edu
Efficacy of Theory-Based Activities for Behavioral Symptoms of Dementia

Ann M. Kolanowski ▼ Mark Litaker ▼ Linda Buettner

Background: Agitation and passivity are behavioral symptoms exhibited by 90% of nursing home residents with dementia. They account for many poor health outcomes, caregiver burden, and increased costs of long-term care.

Objectives: This study tested the efficacy of recreational activities derived from the Need-driven Dementia-compromised Behavior (NDB) model: activities matched to skill level only; activities matched to style of interest only; and a combination of both (NDB-derived) for responding to the behavioral symptoms of dementia.

Methods: Thirty participants were randomly assigned to 1 of 6 possible order-of-condition presentations in this crossover experimental design with repeated measures of dependent variables. Trained research assistants, blind to condition match, implemented each condition for 12 consecutive days. Measures of engagement (time on task and participation), affect, and behavioral symptoms (agitation and passivity) were taken from videotape recordings of each session. Mood was measured with the Dementia Mood Picture Test. The primary analysis method was mixed-model analysis of variance.

Results: Significantly more time on task, greater participation, more positive affect, and less passivity were found under NDB-derived and matched to interest only treatments compared with the matched to skill level only treatment or baseline. Agitation and negative affect improved under all treatments compared with baseline. There was no significant change in mood.

Discussion: The NDB-derived activities are tailored to meet individual needs and improve behavioral symptoms associated with dementia. These findings help to explain factors that produce behavioral symptoms and the mechanisms that underlie their successful treatment.

Key Words: activity interventions · behavioral symptoms · dementia

Agitation and passivity are behavioral symptoms exhibited by 90% of nursing home (NH) residents with dementia and account for many poor health outcomes, including decline in physical functioning, social isolation, and increased risk of abuse (Cohen-Mansfield, Marx, & Rosenthal, 1989; Dyer, Pavlik, Murphy, & Hyman, 2000; Galynker, Roane, Miner, Feinberg, & Watts, 1995; Harwood, Barker, Ownby, & Ducra, 2000). These symptoms have contributed significantly to long-term care costs and have been a major source of caregiver burden (Donaldson, Tarrier, & Burns, 1997; Murman et al., 2002). As dementia progresses, many individuals exhibit both agitated and passive behaviors (Rubin, Morris, & Berg, 1987). This makes their pharmacological treatment difficult because the sedative effects of drugs used to treat agitation may increase passivity. Nonpharmacological interventions have been recommended as the first line of treatment for the behavioral symptoms of dementia (Teri et al., 2002).

Nursing science has few effective interventions for managing behavioral symptoms of dementia because these interventions have lacked a comprehensive theoretical base that takes the root causes into account. Theory-based interventions effectively target treatments. The purpose of this study was to test the efficacy of recreational activities derived from the Need-driven Dementia-compromised Behavior (NDB) model for responding to the behavioral symptoms of agitation and passivity in NH residents with dementia.
Agitation is (a) defined as verbal, vocal, or motor activity that may be abusive or aggressive toward self or others, (b) performed with inappropriate frequency, or (c) considered to be inappropriate by caregivers according to social standards for the specific situation (Cohen-Mansfield et al., 1989). Passivity is characterized by a lessening of mental processes, a decrease in ability to experience or respond to human emotions, fewer interactions with others or the environment, and a decrease in activity (Colling, 2000). Agitation and passivity seem to be opposites, but their causes may be similar: lack of appropriate stimulation from the physical and social environment. A relationship between personal care interactions with NH staff and resident agitation has been reported, especially during bathing (Roth, Stevens, Burgio, & Burgio, 2002; Sloane et al., 1998). Aside from personal care activities, NH residents spend much of their time “doing nothing,” and both agitation and passivity have been observed during these unoccupied times (Cohen-Mansfield, Werner, & Marx, 1992; Logsdon, 2000; MacRae, Schnelle, Simmons, & Ouslander, 1996; Perrin, 1997). Recreational activities are used to fill unoccupied time and may manage behavioral symptoms, but results of efficacy studies were modest (Beck et al., 2002; Opie, Rosewarne, & O’Connor, 1999). A limitation of these studies was that many lacked a theoretical basis for activity prescription.

Behavioral symptoms of dementia are addressed by the NDB model, which is a mid-range theory. The model, published elsewhere (Algase et al., 1996), changes the negative view of behavioral symptoms as “disruptive” or “inappropriate” to a perspective that conceptualizes these behaviors as indicating unmet needs that, if responded to appropriately, will enhance the quality of life. In the model, both background and proximal factors play a role in the occurrence of behavioral symptoms. Background factors are the more stable or slowly changing characteristics of the person with dementia such as neurological factors, cognitive abilities, health status and physical functioning, and psychosocial factors, including premorbid personality. Proximal factors are the more changeable characteristics of the person with dementia and the immediate environment such as physiological and psychological need states and characteristics of the physical and social environment (Figure 1). Some background factors may have a direct influence on behavioral symptoms, independent of proximal factors. Background factors also mediate the response to proximal factors to produce behavioral symptoms, the most integrated response a person can make, given the limitations imposed by the dementia, strengths preserved from abilities and premorbid personality, and the constraints or supports offered by the environment.

Recreational activities derived from the NDB model function as proximal factors that meet individual needs because they are tailored to enrich the physical and social environment by matching to the individual’s background factors. First, NDB-derived activities are matched to the resident’s current cognitive and physical functioning ability so that they are appropriate for his or her level of skill. Skill-appropriate activities not only facilitate engagement but also studies have shown that when people are absorbed in activities that match their ability, they experience positive emotions (Csikszentmihalyi & LeFevre, 1989). Second, NDB-derived activities match interests. The identification of recreational interests in the cognitively impaired is difficult and imprecise, and often done in a trial-and-error fashion. Using the NDB model, the identification of interests is accomplished by a systematic evaluation of premorbid personality, a background factor that can identify style of interest, a lifelong preference for certain types of activities.

### FIGURE 1. Need-driven Dementia-compromised Behavior Model.

<table>
<thead>
<tr>
<th>Background factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurological factors</td>
</tr>
<tr>
<td>Cognitive ability</td>
</tr>
<tr>
<td>Health status (physical functional ability)</td>
</tr>
<tr>
<td>Psychosocial (premorbid personality)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proximal factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physiological need states</td>
</tr>
<tr>
<td>Psychological need states</td>
</tr>
<tr>
<td>Physical environment</td>
</tr>
<tr>
<td>Social environment</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Behavioral symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agitation</td>
</tr>
<tr>
<td>Passivity</td>
</tr>
</tbody>
</table>
A style of interest is defined by the personality traits of extraversion and openness (Costa, McCrae, & Holland, 1984; Holland, 1999). Extraversion reflects the amount of social stimulation preferred by the individual. Persons who rank high on this trait are outgoing and enjoy socializing with others, whereas persons who rank low on this trait prefer more solitary activities. The openness trait reflects the individual's tolerance for the unfamiliar. Persons who rank high on this trait enjoy new activities and like to explore their environment, whereas persons who rank low on this trait prefer more conventional activities. Traits remain relatively stable in adulthood (Hooker & McAdams, 2003), and there is evidence that facets of extraversion and the trait of openness maintain both rank order and mean level stability in dementia (Chatterjee, Strauss, Smyth, & Whitehouse, 1992; Siegler et al., 1991; Strauss, Lee, & Di Filippo, 1997). These long-standing tendencies have been used to identify activity interests in persons with dementia, and improved prescriptive precision over current approaches (Kolanowski, Buettner, Costa, & Litaker, 2001). Personality-based activities are designed to meet individual needs and preferences and thereby reduce behavioral symptoms.

The causal model that underlies the treatment effect is illustrated in Figure 2. It is hypothesized that under implementation of NDB-derived recreational activities, NH residents with dementia will

1. exhibit greater engagement
2. exhibit more positive affect and less negative affect
3. report more positive mood
4. exhibit less agitation and less passivity

To test these hypotheses, (a) time on task and participation (engagement), (b) affect, (c) mood, and (d) behavioral symptoms (agitation and passivity) during implementation of NDB-derived activities were compared with these same variables during baseline and two treatments: activities matched to skill level only and activities matched to style of interest only.

**Methods**

This study used a crossover experimental design with repeated measures of the dependent variables. Measures of engagement, affect, and behavioral symptoms were taken from videotapes using a standard videotape-recording protocol. Measures of mood were obtained in real time by participant interview. The study examined components of the NDB-derived treatment (skill level match and interest match) that were hypothesized to result in therapeutic outcomes. Thirty participants served as their own controls and were assigned by the second author (M.L.) to one of six possible order-of-condition presentations using a permuted blocked randomization scheme. Five participants were assigned to each order of presentation. Trained research assistants, blind to condition match, implemented each activity condition for up to 20 minutes per day for 12 consecutive days, with a 2-day washout period between conditions.

**Participants and Setting**

Effect size estimates were calculated from a pilot study that had a similar design and dependent variables as this study. The power estimates were based on two-sided testing at $\alpha = .05$, with a sample size of 30 participants. Estimated power to detect a medium effect size, expressed as a percentage, for dependent variables were time on task (83%); participation (99%); positive affect (99%); negative affect (40%); mood (66%); and agitation (96%).

The university institutional review board approved the study protocol. Elderly residents were recruited from four NHs located in northeast and central Pennsylvania. Written consent was obtained from each participant’s responsible party and assent from the participant. Participants met the following inclusion criteria: (a) English speaking; (b) diagnosis of dementia that met Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) criteria; (c) had a Mini-Mental State Examination (MMSE;
Behavioral Symptoms of Dementia


Abstract

Background: The study was designed to test the hypothesis that extraversion and openness would be associated with improved engagement in social stimulation and physical activity in NH residents. The study also sought to determine whether a behavioral classification system could be developed and used to identify style of interest.

Methods: We used an activity screen developed for NH residents and a Likert-type self-report for informant use. The NEO-FFI assesses adult personality in five domains: neuroticism, extraversion, openness, agreeableness, and conscionousness. Seven items on hearing, vision, speech, mobility, dressing, personal hygiene, and toileting are rated on a Likert-type scale. Scores range from 0 to 34, with higher scores indicating greater dependency. The physical capacity subscale of the PGDRS has an interrater reliability of .87 and convergent validity with independent measures of nursing time demanded (r = .72; Wilkinson & Graham-White, 1980).

Results: The study found that extraversion and openness were associated with improved engagement in social stimulation and physical activity in NH residents. The behavioral classification system could be developed and used to identify style of interest.

Conclusion: The study suggests that extraversion and openness may be important factors in improving engagement in social stimulation and physical activity in NH residents. The behavioral classification system developed in the study could be useful in identifying style of interest.

Keywords: Extraversion, Openness, Engagement, Social Stimulation, Physical Activity.
The observational scale has descriptive indicators for six affective states: pleasure, anger, anxiety, depression, interest, and contentment. The rater estimated the portion of a 20-minute behavior stream on which any of these affects were evidenced. Scores were obtained for both positive and negative affects, with higher scores indicating more of either affect. Interrater reliability for the ARS was found to be .93.

Mood was measured in real time using the Dementia Mood Picture Test (DMPT; Tappen & Barry, 1995), an instrument that measures both positive and negative moods from the perspective of the cognitively impaired participant. Measures were taken immediately before and immediately after each observation period. The dependent variable for change in mood was the difference between the postmeasurements and premeasurements of DMPT (total) within each day. The participant was shown six "faces" and asked to indicate if the drawing represents how he or she feels. The participant received a total score between 0 and 12, with higher scores representing more positive mood. The instrument has demonstrated high interrater reliability (ICC) of .95 to .99.

Behavioral symptoms were measured using the Cohen-Mansfield Agitation Inventory (CMAI; Cohen-Mansfield et al., 1989) and the Passivity in Dementia Scale (PDS; Colling, 2000). The CMAI is a caregiver-rating questionnaire that consists of 29 agitated behaviors that are rated on a 7-point scale of frequency. The CMAI, modified for direct observation, was used to rate agitation during observation periods (Chrisman, Tabar, Whall, & Booth, 1991). The rater indicated which of the 29 dementia behaviors occurred in 5-minute blocks of time. A sum score was obtained. Interrater reliabilities for the CMAI have ranged from .92 to .95; the scale has reported convergent validity with the Ward Behavior Inventory (Cohen-Mansfield & Billig, 1986).

The PDS is an observer rating scale of 42 behaviors: 12 passive behavior items scored in the negative and 30 active behavior items scored in the positive. Lower scores indicate greater passivity. Five subscale scores were obtained on cognition, emotions, interaction with the environment, interaction with persons, and psychomotor activity. The rater indicated which of the 42 behaviors occurred in 5-minute blocks of time. A sum score was obtained. Internal consistencies (Cronbach’s) of .71 to .94.
were obtained for the subscales and interrater reliability of .80 for the total scale.

Procedure

Prebaseline: Observations made hourly (7 a.m. to 7 p.m.) for 5 consecutive days using the CMAI and PDS included those participants who met enrollment criteria. This was done to determine the type of behavior exhibited and the time of day when these behaviors were most likely to peak.

Baseline: For 12 consecutive days, each participant was observed and videotaped for 20 minutes each day at the time point when he or she exhibited peak behavioral symptoms as determined in prebaseline. If a participant exhibited behaviors at a constant rate across the 7 a.m. to 7 p.m. time frame, or they exhibited several peak times, or they exhibited both agitation and passivity, we randomly selected one of these times/behaviors for observation. Measures of affect and behavioral symptoms (agitation and passivity) were taken from videotapes of each observation session. Mood was measured in real time at the beginning and completion of each observational session by participant interview.

Treatments: The first (A.K.) and third (L.B.) authors prescribed activities based on each participant’s cognitive abilities, physical functioning, and style of interest as assessed by the MMSE, PGDRS, and NEO-FFI, respectively. Each participant’s performance on the individual items that composed the first two instruments was reviewed to determine the cognitive and physical functioning skills they retained. Each participant’s scores on the extraversion and openness scales of the NEO-FFI were used to identify the style of interest category that characterized them. The information on skill level and style of interest guided decisions about activity prescription. Activities were matched to skill level only (treatment A), style of interest only (treatment B), and skill level and style of interest (treatment C). For treatment A, participants received activities matched to their cognitive and physical functioning (skill) level, but opposite their identified style of interest (i.e., from the style of interest category diagonal to their identified style of interest category; Table 1) For example, a participant who scored low on extraversion and low on openness (E−O−) would receive skill-appropriate activities that appealed to artistic interests in a small group (E+O+) for his or her treatment A. For treatment B, participants received activities that matched their style of interest, but not their skill level. For treatment C, participants received activities that were matched to both skill level and style of interest. The research assistants who implemented conditions were undergraduate nursing or recreational therapy students who completed a 2-day training session on the intervention protocol. These trained interventionists were blind to condition match and implemented treatments for up to 20 minutes at each session. Each treatment was given for 12 consecutive days at peak behavior time. To ensure treatment fidelity, random manipulation checks were preformed by the principal investigator on at least 20% of sessions for each condition. These manipulation checks verified, by direct observation, that the interventionist implemented the activity that was prescribed for that session and maintained the activity protocol. If the protocol was not maintained, re-training occurred.

Statistical Methods

Data analysis was carried out according to a preestablished “on treatment” analysis plan that included all observations obtained for each of the 30 participants who had at least partial data for baseline and all three treatment conditions. Sample distributions were examined for each variable, within participant and treatment. Measured values were plotted by day of treatment for each participant to evaluate possible trends across the days of observation for each treatment condition. Distributions of residual values were examined for the normal-distribution-based statistical models to ensure that the assumptions of the analytic methods were met. The primary analysis method was mixed-model analysis of variance (ANOVA) using participant as the random effect and treatment as a fixed effect. Post hoc pair-wise comparison of treatment means was performed using Tukey’s test and was based on least squares means to account for unequal replication due to missing values. Comparison of each treatment mean versus the mean for the baseline condition was performed regardless of the results of the overall test for differences among treatment means. As these were preplanned pair-wise comparisons, each was performed at the 95% confidence level (CI). The analysis was done in two steps. First, separate ANOVA analyses were performed for each of the treatments to evaluate change across days for time on task, positive affect, negative affect, mood, agitation, and passivity. The analysis of participation scores was analogous, but used generalized estimating equation analysis with a multinomial model for the four possible values of the participation score (0, 1, 2, 3). Following this analysis, each participant’s scores were averaged across days within treatments to address large differences in within-participant variability, and mixed-model ANOVA was then used to compare mean scores among the treatments.

Results

The means, standard deviations, and 95% CIs for the dependent variables are listed in Table 2. No significant trend across days of treatment was found for any of the dependent variables. There was a significant difference in mean time on task among the treatments (p = .001). The least squares means for treatment C was significantly higher than for treatment A (p = .001), but not significantly different from treatment B (p = .371). Treatment B was significantly higher than treatment A (p = .040). There was a significant difference in mean participation among the treatments (p < .001). The least squares means for treatment C was significantly higher than that for treatment A (p < .001) or treatment B (p = .003). Treatments A and B were not significantly different (p = .442).

There was a significant difference in positive affect among the treatment means (p < .001). Positive affect was significantly lower for baseline than for treatments B (p = .009) or C (p < .001), but not for treatment A (p = .124). Positive affect was significantly higher for treatment C than for treatment A (p ≥ .021), but not for treatment B.
(p = .219). Treatments A and B were not significantly different (p = .748). There was not a significant difference among the treatment means for negative affect, although the p was very close to the .05 cut point (p = .056), suggesting there may be some treatment effect. The pre-planned comparisons with baseline do show significant differences for baseline versus treatment A (p = .046), baseline versus treatment B (p = .011), and baseline versus treatment C (p = .042). There was less negative affect under all three treatments compared to baseline.

There was no significant difference in mood change score (post, pre) among treatments (p = .860) and none of the treatments were different from baseline (p = .542 for treatment A, p = .997 for treatment B, p = .831 for treatment C).

For agitation, participants showed little variability in CMAI scores across days within treatments. There was a significant difference in mean score among the treatments (p < .001). Under treatments A, B, and C, there was significantly less agitation (p = .007 for treatment A, p = .001 for treatment B, p = .002 for treatment C) than during baseline. There were no significant differences among treatments A, B, and C (all p > .940).

For each subscale of passivity, all three active treatments significantly reduced passivity compared with baseline, with the exception of emotions, where treatment A did not differ from baseline. The treatment comparisons indicated that for each subscale, treatment C resulted in significantly less passivity compared with treatment A, but not treatment B, and treatments A and B did not differ. Significance levels for each subscale are as follows: (a) for thinking, there was a significant difference between treatment means (p < .001). Treatments A, B, and C were significantly different from baseline (p = .026 for treatment A, p = .002 for treatment B, and p < .001 for treatment C). Treatment C differed significantly from treatment A (p = .033), but not treatment B (p = .220). Treatments A and B were not different (p = .833); (b) for emotions, there was a significant difference between treatment means (p < .001). Baseline was significantly different from treatments B (p = .047) and C (p < .001), but not from treatment A (p = .103). Treatments A and C differed significantly (p = .043), but there was no difference between treatments A and B (p = .987) or B and C (p = .096); (c) for interacting with the environment, there was a
significant difference among treatments (p < .001). All
treatment means differed significantly from baseline (p <
.001 for treatment A, p < .001 for treatment B, and p <
.001 for treatment C). Treatment C differed from treat-
ment A (p = .016), but not B (p = .135). Treatments A and
B were not different (p = .349); (d) for interacting with
persons, there was a significant difference among treat-
ments (p < .001). Baseline was significantly different from
treatments A (p < .001), B (p < .001), and C (p < .001).
Treatment C was different from treatment A (p = .016)
but not B (p = .225). Treatments A and B did not differ
(p = .679); (e) for psychomotor activity, there was a sig-
nificant difference among treatments (p < .001). Baseline
was significantly different from treatments A (p < .001), B
(p < .001), and C (p < .001). Treatment C was different
from treatment A (p = .046), but not B (p = .305). Treat-
ments A and B did not differ (p = .802).

Discussion
In this treatment efficacy study, it was hypothesized that par-
ticipants would exhibit improved outcomes under implemen-
tation of NDB-derived recreational activities that were
matched to their skill (cognitive and physical functioning)
level and style of interest (premorbid personality) as com-
pared with recreational activities matched to only one of
those treatment components or baseline. It was found that
agitation and passivity responded best to different treat-
ments, but that NDB-derived activities were efficacious for a broader
spectrum of behavioral outcomes than either of the other condi-
tions. These outcomes were obtained without occurrence of
any adverse events attributable to the treatments.

It was hypothesized that participants would exhibit
greater engagement (time on task and participation) under
NDB-derived activities than under comparison activities
(matched to skill level only or matched to interest only).
This hypothesis was partially supported. Participants spent
more time on task when the activity captured their interests;
that is, it was tailored to either their interests and skills
(NDB-derived) or interests alone, and thus was consistent
with their personality. However, they were scored (timed)
as being “on task” when they were both actively and pas-
sively engaged. Participation was the measure that differen-
tiated levels of engagement, and they participated more
actively when the activity was tailored to both interests and
skill level (NDB-derived). Interest match may be a key treat-
ment component for maintaining attention, but partici-
pants cannot fully participate in activities that require skills
they have lost. Grant and Potthoff (1997) found that use of
skill-appropriate activities improved participation in their
study of NH residents with dementia. Participation was
improved over either treatment component alone by match-
ing activities to both skill level and interests.

It was hypothesized that participants would exhibit
more positive affect and less negative affect under NDB-
derived activities than under baseline or the comparison
activities. Partial support for this hypothesis was found.
Positive affect was improved over baseline and activities
matched to skill level only, when implemented activities
were matched to interests only or matched to both interest
and skill level (NDB-derived). Like time on task, positive
affect responded best when the participant’s style of interest
was identified and used in the prescription of activities.
Alternatively, a weak treatment effect for negative affect
was found. A preplanned comparison indicated that any
type of activity improved negative affect over baseline. Neg-
avative affect was not frequently observed in these partici-
pants. In addition, a power of 40% was used to detect a
treatment effect for negative affect. These issues most likely
contributed to the lack of significant findings for negative
affect. These findings are similar to Beck et al. (2002) who
found significantly more positive affect, but no reduction in
negative affect or agitation, following implementation of
tailored behavioral interventions for NH residents with
dementia. They too found little negative affect, and con-
cluded that their nontargeted interventions need to be more
precisely designed to improve a broader range of behavioral
outcomes. The current NDB-derived interventions were tar-
geted at unoccupied time, but a larger sample size could
have provided more definitive findings on the relationship
between the intervention and negative affect.

It was hypothesized that participants would report more
positive mood under NDB-derived activities than under base-
line or the comparison activities. This hypothesis was not
supported. No significant change in mood under any of our
treatments was found. However, in pilot work, it was found
that NDB-derived activities improved mood over other active
treatments when activities were given twice a day for 3
weeks, suggesting that dosage may be an important factor
here (Kolanowski, Litaker, & Baumann, 2002). The power
was somewhat low for testing this hypothesis and a larger
sample might yield significant results. These results may stem
from a limitation of the self-report method in this population.

It was hypothesized that participants would exhibit less
agitation and less passivity under NDB-derived activities
to under baseline or the comparison activities. Partial sup-
port for this hypothesis was found. Like negative affect, agi-
tation did not demonstrate a differential response to any of
the active treatments. Any treatment reduced agitation com-
pared with baseline. It may be that the diversion present in
any type of activity is sufficient for the successful treatment
of agitation. However, we did not look at the impact of our
interventions outside of treatment times. Work by Kovach
and Wells (2002) indicates that balancing the daily activity
schedule so that residents are not overaroused or under-
aroused for long periods of time reduces agitation. The
NDB-derived activities are designed to be compatible with
residents’ stimulation needs and may be well-suited to main-
taining arousal balance throughout the day while minimizing
polypharmacy. Evaluation of their efficacy is needed within
the context of longer periods of time in addition to their
immediate effect during treatment. Passivity, on the other
hand, responded best when activities matched interests either
alone or matched to both interest and skills. Passivity is
reported to be particularly resistant to intervention (Everitt,
Fields, Soumerai, & Avorn, 1991), and residents who display
this behavior are more behaviorally activated when activities
are designed to be compatible with their individual needs for
social stimulation and novelty. Because passive residents are
at a high risk for functional decline, nonpharmacological
interventions that reduce withdrawn behavior without trou-
blesome adverse effects are particularly advantageous.
The current findings support the use of the NDB model as a framework for understanding the behavioral symptoms of dementia and help to elucidate the mechanisms that underlie their successful treatment. Studies have shown relationships between the individual NDB background factors of premorbid personality traits, cognitive abilities, and physical functioning and the behavioral symptoms of dementia (Harwood et al., 2000; Kolanowski, Strand, & Whall, 1997; Strauss et al., 1997). The current study confirms the importance of these background factors, especially premorbid personality, when responding to the behavioral symptoms of dementia. Prior research has indicated that recreational activities capture interest when they meet individual needs (Tinsley & Eldredge, 1995). Style of interest, which is based on an assessment of personality, is an individual's long-standing disposition to gratify needs in a particular manner (Costa & McCrae, 1998). The NDB model posits that when proximal factors are manipulated in a way that meets individual needs, behavioral symptoms are reduced. The recreational activities that were tailored to style of interest reduced passivity to a greater extent than activities not tailored to style of interest. In addition, engagement and positive affect were improved under interest-matched activities. On the other hand, negative affect, mood, and agitation may be behavioral symptoms that are not be fully explained by the NDB model as currently conceptualized. On the basis of these findings, agitation and negative affect may not require a prescriptive approach beyond simple diversion. It is noted, however, that dosage, the number of times the activity is given each day and the length of each treatment, may be an issue here. A recent study by Kovack et al. (2004) has shown that reducing arousal imbalances by scheduling activities so that residents were not overaroused or underaroused for long periods of time throughout the day reduced agitation. Research using larger sample sizes or more frequent doses of NDB-derived activities, or both, may help clarify the model's utility for explaining these behavioral outcomes.

Behavioral interventions are recommended as a first line of treatment for the behavioral symptoms of dementia. Because they are tailored to the individual's profile, NDB-derived activities meet individual needs. Thus they have the potential to reduce behaviors that signify unmet needs and to promote behaviors that indicate improved quality of life.

References
Series B Psychology Sciences and Social Sciences, 58(6), P296–P304.
An Analytic Strategy for Modeling Multiple-Item Responses
A Breast Cancer Symptom Example

Ardith Z. Doorenbos ▼ Natalya Verbitsky ▼ Barbara Given ▼ Charles W. Given

**Background:** Item Response Theory (IRT) is increasingly applied in health research to combine information from multiple-item responses. IRT posits that a person’s susceptibility to a symptom is driven by the interaction of the characteristics of the symptom and person. This article describes the statistical background of incorporating IRT into a multi-level framework and extends this approach to longitudinal health outcomes, where the self-report method is used to construct a multi-item scale.

**Methods:** A secondary analysis of data from 2 descriptive longitudinal studies is performed. The data include 21 symptoms reported across time by 350 women with breast cancer. A 3-level hierarchical linear model (HLM) was used for the analysis. Level 1 models the item responses, consisting of symptom presence or absence. Level 2 models the trajectory of each individual, representing change over time of the IRT-created latent variable symptom experience. Level 3 explains that trajectory using person-specific characteristics such as age and location of care. The purpose of the analysis is to examine if older and younger women with breast cancer differ in their symptom experience trajectory after controlling for location of care.

**Results:** Fatigue and pain were the most prevalent symptoms. The symptom experience of women with breast cancer was found to improve over time. Neither age nor location of care was significantly associated with the symptom experience trajectory.

**Discussion:** Embedding IRT into an HLM framework produces several benefits. The example provided demonstrates benefits through the creation of a latent symptom experience variable that can be used either as an outcome or as a covariate in another model, examining the latent symptom experience trajectory and its relationship with covariates at the individual level, and managing symptom nonresponse.

**Key Words:** cancer symptoms • hierarchical linear model • Item Response Theory • women with breast cancer

In studying symptoms of disease, exposure to risk, behavior, beliefs, and attitudes, nursing researchers frequently have to combine a number of item responses. In such studies, participants may be repeatedly assessed over time; nested within social settings, such as hospitals, nursing homes, or communities; or both. With social settings in particular, a wide variability of gathered data results from a sometimes unknown multitude of sources. Missing item-level data are often unavoidable as well. Some examples of questions that arise in such cases, requiring robust analytical methodologies are: How does the symptom experience of women with breast cancer change over the course of the chemotherapy treatment? Does age affect engagement in risky behaviors that contribute to AIDS/HIV? Do individual beliefs and cultural attitudes influence the acceptance of differing end-of-life care paradigms?

**Item Response Theory**
Item Response Theory (IRT) was developed in the 1980s in educational research to address some of the issues of measurement practices in scoring tests (McDonald, 1999; van der Linden & Hambleton, 1997). The IRT models postulate that characteristics of a test item, such as its difficulty, interact with an individual’s ability or trait to determine the probability of a correct response to that item (Cheong & Raudenbush, 2000; Lord, 1980). The simplest IRT model, the Rasch model, has only one parameter per item, namely difficulty. The Rasch model makes the assumption that each item is equally discriminating. When this assumption is true, the resulting scale has a clear interpretation—that difficult items will be answered correctly less frequently than easy items. Besides an item-difficulty scale, the IRT can also provide estimates of latent abilities.

Ardith Z. Doorenbos, PhD, RN, is Postdoctoral Fellow; Barbara Given, PhD, RN, FAAN, is University Distinguished Professor, College of Nursing; and Charles W. Given, PhD, is Professor, Department of Family Practice, College of Human Medicine, Michigan State University, East Lansing.

Natalya Verbitsky is Doctoral Candidate, Department of Statistics, The University of Michigan, Ann Arbor.
that can be studied as either explanatory or outcome variables in other models. Of particular significance to this class of analysis, IRT reduces the skewness that commonly arises in composite measures, such as the sum or proportion. This framework has recently been applied to health outcomes (Fortinsky, Garcia, Sheehan, Madigan, & Tullai-McGuinness, 2003; Hays, Morales, & Reise, 2000).

Hierarchical Linear Models

Statistical models that account for nesting of data (e.g., hierarchical linear models [HLMs]) have been growing in popularity (Raudenbush & Bryk, 2002). There has been an increase in the use of HLM in nursing research, especially in research examining patient and organizational outcomes (Cho, 2003; Cho, Ketefian, Barkauskas, & Smith, 2003; Whitman, Davidson, Sereika, & Rudy, 2001). In the past, when confronted with data on individuals nested in organizations, a researcher had to decide whether to perform the analysis at the individual level, thus ignoring the nested structure of the data, or whether to aggregate the variables to the higher level, thus ignoring individual variation within the organizations. In using an HLM analysis, the researcher no longer has to decide at which level to perform the analysis. This avoids problems of misestimating standard errors and of incorrect statistical inference.

There are several benefits of incorporating the IRT into an HLM framework for nursing researchers: (a) it includes the ability to examine multiple dimensions of abilities, traits, or symptoms; (b) it can separate the variation between social settings, such as hospitals, nursing homes, or communities, from the variation between individuals who are nested within these settings; (c) it provides a way to examine the measurement error in the assessment of social settings where individuals are used as informants about their social setting; (d) it allows the researcher to examine the relationship between explanatory variables at various levels (e.g., individual or setting) and the ability or trait; (e) it provides a framework for incorporating repeated observations of item responses to examine changes in the latent ability over time; and (f) the combined framework also provides a way to manage item nonresponse (Raudenbush & Bryk, 2002). These items exemplify the benefits of embedding an IRT model into an HLM framework as a tool for studying symptoms and other self-reported health behavior. A more detailed theoretical discussion regarding incorporating the IRT into an HLM framework can be found in Raudenbush, Johnson, and Sampson (2003), and Johnson and Raudenbush (in press).

Having described the statistical background of incorporating IRT into an HLM framework, the purpose of this article is to illustrate this methodology using an example of the symptom experience for women with breast cancer. This demonstrates the methodology by extending the approach to longitudinal data with health outcomes, where the self-report method is used to construct a multi-item scale. The aim of the analysis is to examine if older and younger women differ in their symptom experience trajectory after controlling for location of care.

An Example: Three-Level HLM Model Incorporating a Symptom IRT

An important aspect of symptom research is how symptom experience varies over time according to the characteristics of the individual and setting. For example, the symptom experience may change differently over time for each woman with breast cancer. Age may influence the relationship, as older women may tend to report fewer symptoms and thus have better symptom experience than younger women at diagnosis and start of chemotherapy. However, younger women may tend to return to the prediagnosis symptom experience faster than older women. Additionally, medical care can affect the symptom experience trajectory. Women receiving care at urban hospitals may have a greater accessibility to medical treatments and thus experience fewer symptoms overall than those at rural hospitals. Incorporating IRT into an HLM framework allows us to examine these and other similar questions.

This example describes a longitudinal Rasch model, which incorporates repeated measures on 21 symptoms at four time points over a 1-year period. Following HLM terminology, we have symptoms at Level 1 nested in repeated measures at Level 2 that are, in turn, nested in individuals at Level 3. To keep the model simple, only three covariates are included in the model: two individual characteristics (age and location of care) and time since diagnosis. In this analysis, the Rasch model orders the responses to a set of items (symptom presence or absence) according to a symptom’s characteristic of prevalence in lieu of the traditional “item difficulty.” The analog to the typical IRT latent ability is then a latent symptom experience.

Data and Participants

This example involves a secondary analysis of data from two descriptive longitudinal studies conducted from 1990 to 1998. There were 242 women from urban hospitals in the first study; 108 women from rural hospitals participated in the second study. These 350 women were newly diagnosed with breast cancer and undergoing chemotherapy. The participants were followed for 1 year and completed telephone interviews on four occasions. At each interview, the presence of 21 symptoms was recorded along with other characteristics.

Inclusion criteria for the primary studies required that women with breast cancer be at least 21 years of age; cognitively intact; and able to speak, read, and write English. Women under the care of a psychologist or psychiatrist, or with a diagnosed emotional or psychological disorder, were excluded. Nurse recruiters approached women who met the inclusion criteria, explained the studies, and obtained written consent. At mutually convenient times, the participants were interviewed by telephone; they also completed self-administered questionnaires. The ages of the participants...
The simplest IRT model, the Rasch model, has only one parameter per item, namely difficulty.

**Measures**

Symptoms were assessed using the self-report Physical Symptom Experience tool (Given et al., 1993). Participants responded regarding the presence of 21 symptoms commonly experienced by individuals with cancer, indicating whether they experienced the symptom (1) or not (0).

Time was coded in days since diagnosis. Demographic information included age and location of care. Location of care was coded rural = 1 if a rural hospital and rural = 0 if an urban hospital. To render the intercept of the regression line meaningful, age was grand-mean centered.

**Analysis**

**Level 1 Model**

The Level 1 model is a standard one-parameter item response or Rasch model, with random effects. In applying the Rasch model, item difficulty was used as symptom prevalence. Let \( Y_{jk} = 1 \) if the symptom \( i \) was present at time \( j \) for person \( k \) and 0 otherwise. The probability of a symptom being present, \( Pr(Y_{jk} = 1) \), is denoted by \( \mu_{jk} \).

At this level, there are 20 dummy variables, \( D_{mjk} \), representing 20 of the 21 symptoms measured. So the Level 1 equation is

\[
\log \left( \frac{\mu_{jk}}{1 - \mu_{jk}} \right) = \pi_{0jk} + \sum_{m=1}^{20} \pi_{mjk} D_{mjk}
\]

\( \pi_{mjk} \) is interpreted as the prevalence of the symptom \( m \) at time \( j \) for person \( k \), compared with the reference symptom (i.e., the symptom for which a dummy variable was not included in the model). This model creates an interval scale for the symptoms, where large values of \( \pi_{mjk} \) indicate more prevalent symptoms while low values indicate less frequent symptoms. By IRT convention, the prevalence for the reference symptom (fatigue) is fixed at 0. This generates the IRT ordering of symptoms by prevalence.

The IRT latent variable describing aggregated symptoms is symptom experience, or \( \pi_{0jk} \), which indicates the overall symptom experience at time \( j \) for person \( k \). \( \pi_{0jk} \) becomes an outcome at Level 2, where the symptom experience trajectory is examined. Larger values of \( \pi_{0jk} \) indicate a higher relative prevalence of symptoms, while smaller values indicate a lower relative prevalence of symptoms.

**Level 2 Model**

The Level 2 model accounts for variation in symptom experience over time for each woman with breast cancer. Equation 2 models parameters from the Level 1 model, \( \pi_{0jk} \) and \( \pi_{mjk} \). To conform to the Rasch methodology, we fixed the prevalence of each symptom \( \pi_{mjk} \) across time (Level 2) and individuals (Level 3) in the model. This constraint reflects the belief that, given a symptom experience, random samples of women with breast cancer will experience a symptom with the same prevalence. Otherwise, the symptom may be regarded as biased against a subset of women with breast cancer. At Level 2, the symptom experience \( \pi_{0jk} \) is described as a function of time.

\[
\pi_{0jk} = \beta_{00k} + \beta_{01k} \times \text{time}_{jk} + \nu_{0jk}
\]

\( \beta_{00k} \) and \( \beta_{01k} \) represent the initial symptom experience, and the linear daily rate of change in symptom experience for individual \( k \), respectively. The random effects, \( \nu_{0jk} \), are the deviations at time \( j \) of individual \( k \)'s symptom experience from the predicted. \( \beta_{mjk} \) represents the prevalence of the symptom \( m \), compared with the reference symptom for individual \( k \).

**Level 3 Model**

At Level 3, the symptom experience trajectory is explained using person-specific characteristics, such as age and location of care.

\[
\beta_{00k} = \gamma_{00} + \gamma_{001} \times (\text{age}_k - 67.72) + \gamma_{002} \times \text{rural}_k + \gamma_{003} \times \text{urban}_k
\]

\[
\beta_{01k} = \gamma_{010} + \gamma_{011} \times (\text{age}_k - 67.72) + \gamma_{012} \times \text{rural}_k + \gamma_{013} \times \text{urban}_k
\]

\( \beta_{m0k} = \gamma_{m00}, \) for \( m = 1, \ldots, 20 \)

In this equation, \( \gamma_{00} \) is the expected initial symptom experience for women with breast cancer who are receiving care in urban hospitals; \( \gamma_{010} \) is the expected difference in the initial symptom experience between two women who differ by 1 year in age; \( \gamma_{002} \) is the expected difference in the average initial symptom experience between the rural and urban locations of care; \( \gamma_{011} \) is the expected average daily rate of change in symptom experience for women receiving care at the urban hospitals; \( \gamma_{013} \) is the difference in the expected daily rate of change in symptom experience between two women who differ by 1 year in age; \( \gamma_{012} \) is the difference in the expected daily rate of change in symptom experience between the rural and urban studies; and \( \gamma_{m00} \) is the prevalence of the symptom \( m \) compared with the reference symptom.

By combining the Levels 1, 2, and 3 models, the hierarchical generalized linear model can be estimated. The combined model tests how the log-odds of experiencing a symptom vary with time- and person-specific characteristics, such as age and location of care. This three-level hierarchical model can be viewed as an item–response model embedded within a hierarchical structure, in which repeated measures are nested within women with breast cancer.

Missing data were addressed by using complete case analyses (Little & Rubin, 2002); that is, if at least one symptom was recorded as present or absent during an interview, then this interview information was used in the analysis. The number of symptoms recorded (present or absent) during interviews ranged from 10 to 21. Some women had fewer than four interviews, resulting in 1,184 interviews (rather than \( 350 \times 4 = 1,400 \)) included at Level 2. Since every woman had at least one interview, and all women had both age and location of care recorded, 350 individuals were included in the analysis at Level 3.
Three-Level HLM Model

Nursing Research July/August 2005 Vol 54, No 4

We present two models, an unconditional and a conditional, estimated using HLM 6.20 (Raudenbush, Bryk, Cheong, & Congdon, 2004). The unconditional model has no covariates at Level 2 or 3, which yields a readily interpretable ordering of symptoms as well as the unadjusted symptom experience estimates for each person at each occasion that they recorded presence or absence of at least one symptom. We examined the symptom experience over time by testing linear and quadratic trajectories at Level 2. To test the associations between individual variables and symptom experience, these variables were incorporated into the multivariate model at Level 3. In the final model, Level 1 remains the same as in the unconditional model (see Equation 1), but now, entered into the model are the time-level variable (days since diagnosis) at Level 2, and the individual-level variable (age and location of care) at Level 3, as previously shown in Equations 2 and 3.

**Model Results**

**Unconditional Model**

Fatigue, the most common symptom in the raw data, was used as the reference symptom (Table 1). The results of the unconditional model yield a readily interpretable ordering of symptoms. Figure 1 shows the symptoms organized by their prevalence ($\gamma_{m00}$); the more prevalent symptoms appear at the top (high values), while the less prevalent appear at the bottom (low values). Symptoms appearing close together in Figure 1 have similar symptom prevalence. Construct validity for this scale was confirmed by the fact that pain and fatigue occurred with greatest frequency, which has been well-established in the cancer literature (Given, Given, Azzouz, Kozachik, & Stommel, 2001; Mock, 2003; Patrick et al., 2003). The lowest frequency symptom during chemotherapy treatment for breast cancer was dehydration (Table 1).

This model produced an unadjusted symptom experience estimate for each individual and each occasion when the presence of at least one symptom was recorded (Figure 2). These symptom experience estimates are approximately normally distributed and may be used as either a covariate or an outcome in other models.

**Table 1. Raw Data Symptom Frequency (All Observations Combined)**

<table>
<thead>
<tr>
<th>Symptoms</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
<th>Yes (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatigue</td>
<td>693</td>
<td>491</td>
<td>1,184</td>
<td>58.53</td>
</tr>
<tr>
<td>Pain</td>
<td>462</td>
<td>722</td>
<td>1,184</td>
<td>39.02</td>
</tr>
<tr>
<td>Insomnia</td>
<td>426</td>
<td>758</td>
<td>1,184</td>
<td>35.98</td>
</tr>
<tr>
<td>Dry mouth</td>
<td>402</td>
<td>781</td>
<td>1,183</td>
<td>33.98</td>
</tr>
<tr>
<td>Loss of feeling</td>
<td>326</td>
<td>857</td>
<td>1,183</td>
<td>27.56</td>
</tr>
<tr>
<td>Urinary frequency</td>
<td>301</td>
<td>882</td>
<td>1,183</td>
<td>25.44</td>
</tr>
<tr>
<td>Weakness</td>
<td>289</td>
<td>894</td>
<td>1,183</td>
<td>24.43</td>
</tr>
<tr>
<td>Cough</td>
<td>262</td>
<td>921</td>
<td>1,183</td>
<td>22.15</td>
</tr>
<tr>
<td>Constipation</td>
<td>225</td>
<td>958</td>
<td>1,183</td>
<td>19.02</td>
</tr>
<tr>
<td>Nausea</td>
<td>208</td>
<td>975</td>
<td>1,183</td>
<td>17.58</td>
</tr>
<tr>
<td>Concentration</td>
<td>187</td>
<td>997</td>
<td>1,184</td>
<td>15.79</td>
</tr>
<tr>
<td>Poor appetite</td>
<td>182</td>
<td>1000</td>
<td>1,182</td>
<td>15.40</td>
</tr>
<tr>
<td>Shortness of breath</td>
<td>169</td>
<td>1015</td>
<td>1,184</td>
<td>14.27</td>
</tr>
<tr>
<td>Weight loss</td>
<td>161</td>
<td>1017</td>
<td>1,178</td>
<td>13.67</td>
</tr>
<tr>
<td>Coordination problems</td>
<td>103</td>
<td>1081</td>
<td>1,184</td>
<td>8.70</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>102</td>
<td>1082</td>
<td>1,184</td>
<td>8.61</td>
</tr>
<tr>
<td>Mouth sores</td>
<td>66</td>
<td>1117</td>
<td>1,183</td>
<td>5.58</td>
</tr>
<tr>
<td>Difficulty swallowing</td>
<td>57</td>
<td>1124</td>
<td>1,181</td>
<td>4.83</td>
</tr>
<tr>
<td>Vomiting</td>
<td>46</td>
<td>1138</td>
<td>1,184</td>
<td>3.89</td>
</tr>
<tr>
<td>Fever</td>
<td>45</td>
<td>1138</td>
<td>1,183</td>
<td>3.80</td>
</tr>
<tr>
<td>Dehydration</td>
<td>26</td>
<td>1156</td>
<td>1,182</td>
<td>2.20</td>
</tr>
</tbody>
</table>

Note. Some of the 350 women had fewer than four interviews and the presence of fewer than 21 symptoms recorded at each interview resulting in 1,184 interviews and varying totals.

**Figure 1.** Symptom prevalences ($\gamma_{m00}$) according to the unconditional model.
Conditional Model

Since our focus was to determine the trajectory of symptom experience as well as its association with important individual variables, partial output for Levels 2 and 3 is reported in Table 2. Both the linear and quadratic trajectories were tested; however, the quadratic term did not significantly improve the model fit. Therefore, a linear trajectory was used in the final model. Time was significantly, negatively associated with symptom experience (\( \hat{\beta}_{10} = -.002, p < .001 \)). So, as women with breast cancer moved through the year, on average their symptom experience improved. Since we hypothesized that the symptom experience trajectory may differ according to women’s age and location of care, age and rural were used to explain initial symptom experience and change in symptom experience. Neither age nor location of care was found to be statistically significantly associated with change of symptom experience over time (\( p = .085 \) and \( p = .819 \), respectively), nor were they found to be statistically significantly associated with symptom experience (\( p = .173 \) and \( p = .150 \), respectively; Table 2).

**Discussion**

The benefits of using IRT are illustrated by the results of the unconditional model. First, IRT created a meaningful metric that reflects the varying prevalence of symptoms in women with breast cancer (Figure 1), while reducing the skewness that commonly arises in composite measures of symptoms (Figures 2 and 3). Second, the analysis provided estimates of the latent symptom experience for each person at each occasion when the presence or absence of at least one symptom was recorded. These symptom experience estimates can be used as explanatory or outcome variables in other models.

Several of the numerous benefits of embedding IRT into an HLM framework were illustrated in the example above. First, this methodology provided a framework for incorporating repeated observations on the presence of symptoms.

![FIGURE 2. Histogram of symptom experience from the unconditional model. Some of the 350 women had fewer than four interviews, resulting in 1,184 total interviews.](image1)

![FIGURE 3. Histogram of total number of symptoms present. Some of the 350 women had fewer than four interviews, resulting in 1,184 total interviews.](image2)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept, ( \gamma_{00} )</td>
<td>.605</td>
<td>0.106</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Age, ( \gamma_{01} )</td>
<td>-.010</td>
<td>0.007</td>
<td>.173</td>
</tr>
<tr>
<td>Rural, ( \gamma_{02} )</td>
<td>.268</td>
<td>0.185</td>
<td>.150</td>
</tr>
<tr>
<td>Time, ( \gamma_{10} )</td>
<td>-.002</td>
<td>0.0003</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Time ( \times ) Age, ( \gamma_{11} )</td>
<td>.00005</td>
<td>0.00003</td>
<td>.085</td>
</tr>
<tr>
<td>Time ( \times ) Rural, ( \gamma_{12} )</td>
<td>.0002</td>
<td>0.0007</td>
<td>.819</td>
</tr>
</tbody>
</table>

![TABLE 2. Partial Output of Estimates for the Conditional Model (Excluding Symptoms Prevalence)](table2)
symptoms to examine changes in the latent variable symptom experience over time. The results on the conditional model showed that the symptom experience of women with breast cancer improved over time. Second, this methodology allowed us to examine the relationship between individual variables and the latent symptom experience. Contrary to our hypothesis, controlling for the location of care, no statistically significant association of age with the symptom experience trajectory was found.

Moreover, the hierarchical framework provides a way to manage item nonresponse. Presently, two common approaches for combining symptom information use the sum or the proportion of the symptoms present. When using a summary score, a nursing researcher must make an arbitrary decision regarding how to handle item nonresponse. Using the sum assumes that everyone has the same number of symptoms recorded, but not necessarily present; the symptoms that are not recorded are assumed to be not present. The proportion approach assumes that each symptom contributes the same amount of information, which is again problematic, since some symptoms occur more frequently than others. A researcher can use an IRT model without having to decide what to do with missing data, as long as the data are assumed missing at random, a comparatively mild assumption (Little & Rubin, 2002).

The task of combining information from multiple-item responses arises frequently in studies of health outcomes. In many of these studies, the items are measured over time and nested within individuals, and item-level missing data are often unavoidable. This report demonstrates how embedding a Rasch model into HLM can address these research challenges.


References

Dating Violence in College Women
Associated Physical Injury, Healthcare Usage, and Mental Health Symptoms

Angela Frederick Amar • Susan Gennaro

Background: College-aged women report experiencing violence from a partner within the dating experience.

Objectives: This study used a correlational design, to report physical injury, mental health symptoms, and healthcare associated with violence in the dating experiences of college women.

Methods: A convenience sample of 863 college women between 18 and 25 years of age from a private, historically Black university in the South, and a private college in the mid-Atlantic completed the Abuse Assessment Screen, a physical injury checklist, and the Symptom Checklist—R-90. Data analysis consisted of frequencies, ANOVA, and MANOVA.

Results: Almost half (48%) \((n = 412)\) reported violence and, of these, 39% \((n = 160)\) reported more than one form of violence. The most commonly reported injuries were scratches, bruises, welts, black eyes, swelling, or busted lip; and sore muscles, sprains, or pulls. Victims had significantly higher scores on depression, anxiety, somatization, interpersonal sensitivity, hostility, and global severity index than nonvictims. Victims of multiple forms of violence had significantly higher mental health scores and reported greater numbers of injuries than victims of a single form of violence. Less than half of those injured sought healthcare for injuries and less than 3% saw a mental health professional.

Discussion: Study findings suggest the importance of screening and identification of victims of violence. Knowledge of physical and mental health effects of violence can guide intervention, prevention, and health promotion strategies. Future research is needed to describe barriers to seeking healthcare, screening practices of college health programs, and programs to identify victims.

Key Words: college women • interpersonal violence • mental health symptoms • physical injury

Although dating represents a carefree period of romantic experimentation, for many dating becomes harmful owing to the experience of violence. Prevalence of dating violence ranges from about 30% for physical violence, 8% for stalking, 90% for emotional violence, and 20% for sexual violence (Fisher, Cullen, & Turner, 2000; Johnson & Sigler, 2000; Riggs & O’Leary, 1996; Tjaden & Thoennes, 1998b). Dating violence is the term often used to describe adolescent and college student intimate partner violence.

Intimate partner violence is a pattern of purposeful coercive behaviors that may include inflicted physical injury, psychological abuse, sexual assault, progressive social isolation, stalking, deprivation, intimidation and threats. These behaviors are perpetrated by someone who is, was or wishes to be involved in an intimate or dating relationship with an adult or adolescent victim and are aimed at establishing control of one partner over the other (Family Violence Prevention Fund, 1999).

According to the United States Department of Justice, women aged 16–24 are most at risk for nonfatal violence from an intimate partner (U. S. Department of Justice, 1998). The purpose of this study was to explore the violence that occurs in dating experiences of college women. The specific aims were to: (a) compare mental health symptoms of women who have been a victim of dating violence with those who have not; (b) describe the types of physical injuries resulting from violence in dating experiences and the healthcare sought; and (c) compare mental health symptoms and the number of physical injuries of women who have experienced multiple forms of dating violence.

Angela Frederick Amar, PhD, RN, is Assistant Professor, School of Nursing and Health Studies, Georgetown University, Washington, DC.
Susan Gennaro, DSN, RN, FAAN, is Professor, School of Nursing, University of Pennsylvania, Philadelphia.
violence with women who have experienced one form of violence.

Background

Mental Health Symptoms Associated With Dating Violence

Mental health reactions to victimization include depression, anxiety, posttraumatic stress disorder, somatic complaints, and anger (Bohn & Holz, 1996; Campbell, 2002). However, linkages to mental health symptoms must be viewed with caution, as cross-sectional research cannot account for preexisting conditions.

Severe physical dating violence has been associated with suicidal ideation or attempts in adolescent females (Silverman, Raj, Mucci, & Hathaway, 2001). In a study of psychological abuse using a sample of mainly White females, undergraduate victims reported higher levels of hostility than nonvictims did but there were no differences between the two groups in depression, anxiety, and somatization (Pape & Arias, 1995). In the National College Women Sexual Victimization Survey, three in 10 women reported being “injured psychologically” from stalking (Fisher et al., 2000). In another sample of predominantly White females, stalked individuals reported more posttraumatic stress symptoms and had greater severity of mental health symptoms than victims of harassment or controls (Westrup, Fremouw, Thompson, & Lewis, 1999).

Rape is associated with feelings of helplessness, powerlessness, anxiety, and fear; posttraumatic stress disorder; multiple somatic complaints; and genital injuries (Burgess & Holmstrom, 1974; Koss & Cook, 1998). Shapiro and Schwarz (1997) reported that those who reported date rape endorsed more trauma symptoms than those who did not report rape (Shapiro & Schwarz, 1997).

Physical Injury Associated With Dating Violence

There is limited research to describe physical injury from dating violence and resulting healthcare. Women who reported stalking also reported injuries from physical assault by their stalkers, such as swelling, cuts, scratches, bruises, broken teeth, and knife or gunshot wounds (Kohn, Flood, Chase, & McMahon, 2000). Fisher, Cullen, and Turner (2000) reported that in about one in five rape and attempted rape incidents, victims reported injury, most often bruises, black eyes, cuts, scratches, swelling, or chipped teeth.

The most thorough report on abuse is the National Violence against Women Survey (NVAWS), a telephone survey of a nationally representative sample of 8,000 adult women and 8,000 men on physical abuse, sexual victimization, threats and coercion, stalking, and emotional abuse (Tjaden & Thoennes, 1998a, 1998b, 2000). The survey found that in more than one third of all rapes and physical assaults against women by intimates, the victim sustains an injury and that in about one third of injury cases, healthcare was received. While most of the women reported relatively minor injuries (66–73%), such as scratches, bruises, and welts, relatively few women reported more serious types of injuries (2–17%), such as lacerations, broken bones, dislocated joints, head or spinal cord injuries, chipped or broken teeth, or internal injuries (Tjaden & Thoennes, 1998a, 2000). However, the study was framed as a crime survey, which could have limited responses about interpersonal violence.

Research on Occurrence of Multiple Forms of Violence

Most of the dating violence research has focused solely on physical violence, with an emphasis on perpetration of acts (Jackson, 1999; Lewis & Fremouw, 2001). Limited research on dating violence includes all possible forms of physical, sexual, and psychological violence and stalking. The NVAWS provides compelling evidence of different types of victimization, and that indicates many victims may experience more than one type of violence. Women with emotionally abusive partners were more likely to report being raped, physically assaulted, and/or stalked by their partners even when controlling for sociodemographic and relationship variables (Tjaden & Thoennes, 2000). Similarly, victims of stalking from the National College Women Sexual Victimization Study reported that in 15.3% of the incidents, the stalker threatened or attempted to harm them, and in 10.3% of the incidents, the stalker forced or attempted sexual contact (Fisher et al., 2000).

In a study of sheltered battered women who also reported sexual abuse, 99% of the participants experienced at least one physical health symptom and attributed many of the physical health symptoms to the experience of abuse (Eby, Campbell, Sullivan, & Davidson III, 1995). Similarly, Campbell and Soeken (1999) report that physically abused women who also reported sexual abuse had significantly higher scores on negative health symptoms and gynecological symptoms than women who reported only physical abuse. Although these studies report on adult women, both had participants who were between 18 and 24 years of age and can contribute to understanding the impact of intimate partner violence on the health of women.

The study of dating violence has centered most often on physical violence and the perpetration of violence. Research exploring the health-related effects has been limited, and most studies have used middle class and White samples (Jackson, 1999; Lewis & Fremouw, 2001). This study sought to build on the existing knowledge of dating violence by the inclusion of Black participants and the use of a victim perspective to obtain data on associated physical and mental health effects.

Design and Method

Research Setting and Sample

The study took place at two universities, a historically Black private college in the South, and a private college in the mid-Atlantic region. Data collection occurred on both campuses in residence halls and meeting rooms over the 2002–03 academic year.

To be included in this study, the woman must have been between 18 and 25 years of age, dated a male within the past year, and be able to read English. The researchers anticipated similar prevalence as previously reported and sampled to ensure adequate numbers of victims to create subgroups. A sample size of 863 participants was adequate.
to provide individuals who had experienced dating violence. Power analyses were conducted (Cohen, 1988) and all target sample sizes were smaller than the 863 sample obtained.

**Instrumentation**

**Physical, Sexual, and Psychological Violence**

The Abuse Assessment Screen (AAS) measured the occurrence of intimate partner violence within the past year (Soeken, McFarlane, Parker, & Lominack, 1998). All items yielded a yes/no response that coded participants as victims or nonvictims. The item pertaining to violence while pregnant was excluded because pregnancy precluded study participation. Pregnant women were excluded from the study because the study was not designed to assess and manage the additional concerns of pregnancy. The researchers added an item about stalking or harassment, that used the same terminology and phrasing as the items on the AAS.

The Nursing Research Consortium on Violence and Abuse developed the AAS in 1991, and since then researchers have reported effective use with adolescent and young adult samples (Coker, McKeown, et al., 2000; Curry, 1998; Lown & Vega, 2001). The AAS has been effectively used in studies with Blacks (Coker, Smith, McKeown, & King, 2000; Curry, 1998; Dunn & Oths, 2004). Using a test-retest approach and on comparison with similar measures, the AAS was established as a reliable and valid measure to screen for relationship violence (Soeken, McFarlane, Parker, & Lominack, 1998). The reliability for this study was 0.76 using 851 participants.

**Mental Health Symptoms**

The SCL-90-R measured mental health symptoms and general psychological health. Particular subscales of interest were somatization, interpersonal sensitivity, depression, anxiety, and hostility. The global severity index, based on the total scale, reported overall psychological distress. A number of studies have used the SCL-90-R and have shown it to be a reliable instrument with internal consistency coefficients between 0.80 and 0.90, and a valid instrument through correlations with other instruments (Derogatis, 1994). Researchers report effective use of the SCL-90-R with Black participants (Champion, Shain, Piper, & Perdue, 2002; Martin, Kilgallen, Dee, Dawson, & Campbell, 1998). Reliability coefficients for this study were as follows: total scale = 0.97 (n = 703), somatization subscale = 0.84 (n = 842), interpersonal sensitivity = 0.87 (n = 841), depression = 0.90 (n = 830), anxiety = 0.85 (n = 847), and hostility = 0.79 (n = 850).

**Physical Injury**

Any participant who reported experiencing violence indicated if physical injury occurred. The list of injuries was consistent with the literature on interpersonal violence and the National Violence Against Women Survey (Brockmeyer & Sheridan, 1998; Eby et al., 1995; Tjaden & Thoennes, 1998b). They also reported if they visited healthcare providers after injuries.

**Human Subjects Considerations**

Institutional review board approval was obtained at both settings. To ensure confidentiality, the participants and universities were unnamed. Each survey was numbered, and participants used that number to sign the consent form. Each participant received information on the study, educational pamphlets, and community and campus resources. They also received the phone numbers of one of the researchers (who is a certified Psychiatric Clinical Nurse Specialist) and the university counseling offices. All participants were aware that they could discontinue if any emotional distress occurred. No problems were presented during data collection.

**Data Collection**

After receiving approval from campus officials, the researchers approached young women in designated areas on campus to discuss participation in the study. Potential participants were told about the study, its purpose, and what participation entailed. Study participation included completing a pen-and-paper survey. Participants received no incentives for study participation.

As with most research on partner violence, the researchers were concerned about the safety of the participants (Dutton et al., 2003). All signs and e-mails described the study topic as dating and violence. The surveys were completed either at the researcher’s table or in the participant’s room. Any young women who were accompanied by males were not approached about participation. No safety issues emerged during data collection.

**Data Analysis**

The researchers used Statistical Package for Social Scientists 11.0 (SPSS) for data analysis. All statistical analyses were conducted for α = .05 level of significance. Data analysis for comparing victims with nonvictims included the entire sample of 863 participants, with 412 victims and 451 nonvictims. Subsequent analyses included only the 412 victims. The victims were divided into two groups: those who had experienced any one form of violence (n = 252) and those who had experienced any two or more forms of violence (n = 160). The decisions to divide the sample reflected consideration of keeping the group sizes fairly equivalent to meet the assumptions of ANOVA techniques.

The sample was composed of young women between the ages of 18 and 25 (M = 19.3 years, SD = 1.46). There is little consistency which ages determine late adolescence and early adulthood (Berry, 2004). Although many agree that the age range begins at 18, the ending age of the range varies from 21, 22, or 24 (Cutler & Marcus, 1999; Grace, 1998; Neinstein, Juliani, & Shapiro, 1991). This sample was largely in late adolescence because 94% (n = 813) were ages 18–22 (Cutler & Marcus, 1999). The grade-point average ranged from 1.0 to 4.0 (M = 3.17, SD = 0.51). Participants reported an average of three boyfriends, with a range of 0–30 boyfriends. The most often reported dating situations were (a) involved in a relationship with a boyfriend (40%, n = 343) and (b) not currently dating, but dated within the past year (34%, n = 296).
Most participants were single (99%, n = 855) and Black (70.5%, n = 608). Further ethnic breakdown of the sample was: White (17.5%, n = 151), Asian/Pacific Islander (6.1%, n = 53), American Indian (0.1%, n = 1), Hispanic (1.3%, n = 11), and mixed race/multiracial (4.5%, n = 39). The level of household income was evenly distributed, $50,000–$74,999 (24%, n = 201) as the most frequently reported category.

Results

Assessment of Violence
In this study, 48% (n = 412) of the participants had experienced some form of violence within the past year. Victims who experienced any one form of violence were classified as single and those who experienced any two or more forms of violence were designated as multiple. For example, a young woman who reported only experiencing psychological violence was a single victim, while an individual who reported both psychological and physical violence was classified as a multiple victim. The experience of the two forms of violence may or may not have been within the same relationship or by the same perpetrator. Of the 412 women who reported experiencing intimate partner violence, almost 40% (n = 160) had experienced more than one form of violence. No significant differences existed in age, race, and other demographic variables between victims and nonvictims or in those victims who experienced a single form of violence and those who experienced multiple forms of violence.

Mental Health Symptoms and Dating Violence
Correlational analysis was conducted for mental health symptoms and demographic characteristics. Despite significant correlations (p < .01) among the mental health symptoms, each represented a different concept and was included in the analysis. No other significant correlations existed among mental health symptoms and demographic characteristics. There was a difference between victims and nonvictims on mental health symptoms as analyzed using MANOVA (Wilks’ λ = 0.023, f = 7190.83, df = (5, 857), p < .001). The effect size for the MANOVA was 0.26, which indicates a moderate effect (Cohen, 1988). As seen in Table 1, post hoc ANOVA statistics were significant suggesting that dating violence victims had significantly higher mental health symptom scores than nonvictims. Each ANOVA was found to be significant using Bonferroni’s adjustment as the α = .01 protecting against the inflation of the family-wise error rate associated with conducting multiple ANOVAs.

The global severity index, which indicated general psychological distress, was derived from the total scale necessitating a separate ANOVA. The mean scores for victims (M = 59.44, SD = 10.73) were higher than the scores for nonvictims (M = 53.78, SD = 11.08). Findings suggest that victims had significantly higher general psychological distress than nonvictims had (F = 57.89, df = 1, 861, p < .001).

Derogatis (1994) used the Symptom Checklist-90—Revised to screen for psychiatric disorders and to provide a value that indicated “caseness” or risk for a psychiatric disorder and the need for further diagnostic screening. The criterion for caseness was met by 37% of the participants (n = 319). Not being a victim of dating violence significantly reduced the odds of caseness (OR = 0.463, 95% CI = [0.349, 0.614], p < .01), suggesting that nonvictims had lower odds of meeting the criterion for caseness.

### Table 1. Mental Health Symptoms Victims and Nonvictims

<table>
<thead>
<tr>
<th>Mental Health Symptoms</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>df</th>
<th>p</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Somatization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonvictims</td>
<td>451</td>
<td>50.07</td>
<td>10.10</td>
<td>861</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Victims</td>
<td>412</td>
<td>54.26</td>
<td>10.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interpersonal sensitivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonvictims</td>
<td>451</td>
<td>56.35</td>
<td>10.69</td>
<td>861</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Victims</td>
<td>452</td>
<td>60.49</td>
<td>10.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonvictims</td>
<td>451</td>
<td>53.94</td>
<td>10.53</td>
<td>861</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Victims</td>
<td>412</td>
<td>58.51</td>
<td>9.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonvictims</td>
<td>451</td>
<td>50.02</td>
<td>10.50</td>
<td>861</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Victims</td>
<td>412</td>
<td>53.91</td>
<td>11.42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hostility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonvictims</td>
<td>451</td>
<td>52.10</td>
<td>9.53</td>
<td>861</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Victims</td>
<td>412</td>
<td>57.53</td>
<td>10.68</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. ANOVA = Analysis of variance.
Physical Injury Associated With Violence

<table>
<thead>
<tr>
<th>Injury</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any injuries received</td>
<td>132</td>
<td>32.0</td>
</tr>
<tr>
<td>Scratches</td>
<td>62</td>
<td>15.0</td>
</tr>
<tr>
<td>Sore muscles, sprains, strains, or pulls</td>
<td>60</td>
<td>14.6</td>
</tr>
<tr>
<td>Bruises, welts, black eyes, swelling, busted lip</td>
<td>53</td>
<td>12.9</td>
</tr>
<tr>
<td>Genital injury, bleeding genitalia, sore or irritated genitals</td>
<td>39</td>
<td>9.5</td>
</tr>
<tr>
<td>Acquired an STD</td>
<td>36</td>
<td>8.7</td>
</tr>
<tr>
<td>Bite marks, wounds</td>
<td>21</td>
<td>5.1</td>
</tr>
<tr>
<td>Lacerations, knife wounds, cuts</td>
<td>13</td>
<td>3.1</td>
</tr>
<tr>
<td>Broken bones, dislocated joints*</td>
<td>7</td>
<td>1.7</td>
</tr>
<tr>
<td>Knocked unconscious, passed out*</td>
<td>3</td>
<td>0.7</td>
</tr>
<tr>
<td>Head or brain injury*</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>Chipped or knocked out teeth*</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>Spinal cord injuries*</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Perforated or shattered ear drum*</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Gunshot injuries*</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Severe injuries.

compared with victims. Of the victims, 172 (42%) met the criterion for caseness, as compared with nonvictims, where 113 (25%) met the criterion for caseness. About 60% of those who met the criterion for caseness (n = 190) were victims of dating violence. Physical injury after violence was not significantly associated with caseness.

Physical Injury Associated With Violence

Of the 412 participants who reported experiencing violence, almost a third (n = 132) reported physical injury. The most commonly received injuries were scratches, sore muscles, sprains, strains, or pulls, and bruises, welts, black eyes, swelling, or busted lip. Injuries were categorized as severe and less severe (Brockmeyer & Sheridan, 1998; Eby et al., 1995; Tjaden & Thoennes, 1998b). Most of the reported injuries (89%, n = 118) were less severe. About 13% (n = 52) reported severe injuries and 38 individuals reported both less severe and more severe injuries (see Table 2).

About 40% (n = 56) of the participants who reported injury sought healthcare, which was most commonly an outpatient appointment with a healthcare provider (7%, n = 27), student health services (4%, n = 18) or emergency services (4%, n = 18). Other healthcare providers seen are as follows: mental health professional (3%, n = 11), inpatient hospitalization (1%, n = 5), and ambulance/paramedic services (1%, n = 4). In addition, just over half of the victims told someone that they had experienced violence. Of these, friends were most often told (50%) and clergy the least often (>2%). It is interesting to note that counselors were told at rates of less than 6% and families at less than 25%.

Multiple Victims and Mental Health Symptoms

The MANOVA comparison of mental health symptoms in single and multiple victims was significant (Wilks’ λ = 0.96, F = 3.74, df = 5, 406, p < .003) as were post hoc ANOVAs (Table 3). Again, the effect size was moderate (0.21). As seen in Table 3, victims who experienced multiple forms of dating violence had significantly higher mental health scores as compared to victims who experienced one form of violence. Each ANOVA was found to be significant using Bonferroni’s adjustment as the α = .01 protecting against the inflation of the family-wise error rate associated with conducting multiple ANOVAs.

A separate ANOVA for the global severity index revealed that victims with multiple occurrences had significantly greater general psychological distress than single victims (F = 15.70, df = 1, 410, p < .001). The mean scores for multiple victims (M = 57.80, SD = 10.89) were higher than the mean scores for single victims (M = 52.03, SD = 9.98).

Again, 318 participants (37%) met the criterion for “caseness.” Of those who met the criterion for caseness and victim status, 45% (n = 86) were multiple victims. Single victims had decreased odds of experiencing caseness as compared to multiple victims (OR = .61, 95% CI = (.406, .901), p < .01). Of the victims who experienced multiple forms of violence, 78 (49%) met the criteria for caseness, as compared with victims who experienced a single form of violence, among whom 94 (37%) met the criteria for caseness.

Multiple Victims and Number of Injuries

The mean number of injuries reported from single victims was 0.33 (SD = 0.90) and from multiple victims was 1.43 (SD = 1.68). Significant differences were revealed by the ANOVA in the number of injuries reported by each group (F = 74.96, df = 1, 41, p < .01). As one would expect, the number of injuries reported by multiple victims was significantly greater than the number of injuries reported by single victims. However, the relationship between the severity of injuries and the number of forms of violence was not significant.

Discussion

Limitations of the study include representativeness of the sample, premorbid conditions, sampling bias, and self-report concerns. Because the study contained Black college women, results should not be generalized to other groups. One limitation of the retrospective, cross sectional design is that causality and preexisting conditions cannot be taken into account; this limitation applies to mental health symptoms.

In the statistical analyses the assumption was that the sample represented a normally distributed population, but in reality the convenience sample was a nonprobability one, which is a limitation on the generalizability of the findings. Individuals could choose to participate or not because of previous experience with violence, which could lead to inaccurate prevalence rates because of sampling bias. As with any method using self-reports, there was the possibility that participants could have underreported...
abuse or responded to items in what they felt was a positive manner.

The mean scores of victims on all mental health symptoms were significantly higher than the scores of nonvictims, which have been reported in previous research. In this research premorbid conditions and causality are not taken into account, but previous research on domestic violence suggests that mental health sequelae are the likely outcomes of intimate partner violence (Campbell et al., 2002).

Almost one third of the sample met the criteria for psychiatric diagnosis, and further evaluation that young women are not seeking and/or getting mental health evaluation and treatment. This was consistent with prior literature that only a minority of victims reported using counseling or supportive services (Henning & Klesges, 2002). Barriers to disclosure in healthcare settings of adult domestic violence victims have been identified through research. Common barriers include fear of retaliation, embarrassment, lack of treatment or scorn from the provider, misunderstanding from the staff, fear, and a lack of resources (D’Avolio et al., 2001; Yam, 2000). In addition, because women often disclose abuse to friends and not to family members, counselors, or clergy—all of whom could provide assistance otherwise in getting necessary intervention.

Many participants who reported violence also reported physical injury. Of those reporting physical injury, at least a fifth reported injuries that were more serious. This was consistent with the findings of the National Violence against Women Survey, where about one third of rapes and physical assaults by an intimate partner resulted in physical injury and most of the injuries were rated as less severe (Tjaden & Thoennes, 2000). Yet, in this study, less than half of the injured received medical attention for their injuries, most from either outpatient medical appointments or student health services. These findings are not surprising in light of similar findings among adult battered women populations.

Women are often hesitant to report violence (Valente, 2000) and many report not discussing intimate partner violence with a physician (Plichta & Falik, 2001). Again, research that explores perceived barriers to discussing abuse in adolescent populations is limited. These findings underscore the necessity for routine screening of all adult and adolescent women for the experience of intimate partner violence at healthcare setting visits (Groves, Augustyn, Lee, & Sawires, 2002).

Because most of the reported injuries (89%, n = 118) were classified as less severe or minor injuries like scratches, sore muscles, and bruises, these types of injuries could serve as triggers to assess for the existence of a violent dating experience or partner. Because some participants reported sexually transmitted diseases related to a violent dating encounter, it is important for nurses who work with college-aged females to inquire about experiences of violence with individuals who are seen for sexually transmitted diseases. While the presence of these physical injuries does not provide conclusive evidence of intimate partner violence, they can urge the nurse to screen carefully. Further research can determine which screening techniques work best with college women.

The association of mental health symptoms with physical, sexual, and emotional violence, and stalking has been documented (Campbell, 2002; Campbell, Jones, & Dienemann, 2002; Campbell & Soeken, 1999; Golding, 1999). Victims of single forms of violence had a significantly

<table>
<thead>
<tr>
<th>Mental Health Symptoms</th>
<th>Single</th>
<th>Multiple</th>
</tr>
</thead>
<tbody>
<tr>
<td>Somatization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>252</td>
<td>160</td>
</tr>
<tr>
<td>Multiple</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>252</td>
<td>160</td>
</tr>
<tr>
<td>Multiple</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>252</td>
<td>160</td>
</tr>
<tr>
<td>Multiple</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hostility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>252</td>
<td>160</td>
</tr>
<tr>
<td>Multiple</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 3. Mental Health Symptoms Single and Multiple Victims
decreased likelihood of caseness than victims of multiple forms of violence. The experience of violence is a predictor of future violence (Smith, Thornton, DeVellis, Earp, & Coker, 2002). Therefore, young women who have already experienced more than one form of violence are at risk for continued abuse, which could affect long-term mental health. The cross-sectional approach cannot account for premorbid diagnosis, and thus causality cannot be assumed. However, research on adult women makes it quite likely that an association exists between intimate partner violence and mental health symptoms (Campbell, 2002).

As a young woman experienced more forms of violence, the effects of violence on health increased. Victims of multiple forms of violence reported more physical injuries than victims of single instances. The more the violence that one was exposed to, the greater the likelihood of injuries. It seems plausible that young women who experience more than one form of violence could be involved with partners who are more violent. Overall, a partner who perpetrates multiple forms of violence could have behavior that is more violent. Alternatively, individuals who have experienced multiple forms of violence may have had more than one violent relationship that could have increased the number of injuries sustained. Physical injury represents a threat to the physical and mental health of an individual.

It is suggested in this study that young women do not report or disclose the consequences of violence in the healthcare setting. Perceived barriers to seeking treatment, disclosing, and/or reporting intimate partner violence would improve health promotion activities.

Over half of the women in the sample experienced violence suggesting that many college women are susceptible to the health effects associated with dating violence. Mental health symptoms were significantly greater in women who had experienced violence as compared with women who had not. Mental health symptoms and the number of physical injuries also were significantly greater in women who had experienced more than one form of violence as compared with women who had experienced one form of violence. Almost one third of the participants who reported the experience of violence also reported physical injury. Of those reporting physical injury at least, one fifth reported injuries that were more serious, and yet few sought healthcare.

References


The Effectiveness of a Nursing Inpatient Smoking Cessation Program in Individuals With Cardiovascular Disease

Maud-Christine Chouinard ▼ Sylvie Robichaud-Ekstrand

► **Background:** Smoking is an important risk factor for cardiovascular disease (CVD), and quitting is highly beneficial. Yet, less than 30% of CVD patients stop smoking. Relapse-prevention strategies seem most effective when initiated during the exacerbation of the disease.

► **Objective:** A nurse-delivered inpatient smoking cessation program based on the Transtheoretical Model with telephone follow-up tailored to levels of readiness to quit smoking was evaluated on smoking abstinence and progress to ulterior stages of change.

► **Method:** Participants (N = 168) were randomly assigned by cohorts to inpatient counseling with telephone follow-up, inpatient counseling, and usual care. The inpatient intervention consisted of a 1-hr counseling session, and the telephone follow-up included 6 calls during the first 2 months after discharge. The nursing intervention was tailored to the individual’s stage of change. End points at 2 and 6 months included actual and continuous smoking cessation rates (biochemical markers) and increased motivation (progress to ulterior stages of change).

► **Results:** Assuming that surviving patients lost to follow-up were smokers, the 6-month smoking abstinence rate was 41.5% in the inpatient counseling with telephone follow-up group, compared with 30.2% and 20% in the inpatient counseling and usual care groups, respectively (p < .05). Progress to ulterior stages of change was 43.3%, 32.1%, and 18.2%, respectively (p = .02). Stage of change at baseline and intervention predicted smoking status at 6 months.

► **Discussion:** This tailored smoking cessation program with telephone follow-up significantly increased smoking cessation at 6 months, and progression to ulterior stages of change. The telephone follow-up was an important adjunct. It is, therefore, recommended to include such comprehensive smoking cessation programs within hospital settings for individuals with CVD.

► **Key Words:** cardiovascular disease • smoking cessation • transtheoretical model

Smoking has been a major public health problem in the world (Fiore et al., 2000), and has been reported as an important risk factor for cardiovascular disease (CVD; American Association of Cardiovascular and Cardiopulmonary Rehabilitation, 2004; Pipe, 1999). Smoking cessation has been highly beneficial for persons suffering from CVD (Critchley & Capewell, 2003; Wilson, Gibson, Willan, & Cook, 2000), yet less than 30% have stopped smoking, despite advice from health care professionals (van Berkel, Boersma, Roos-Hesselink, Erdman, & Simoons, 1999).

Several reports indicate that the exacerbation of a disease motivates individuals to stop smoking (Richmond, 1999), and that the hospital smoke-free environment is an important contributing factor (Rigotti, 2000). Taylor, Houston-Miller, Killen, and DeBusk (1990) reported a 61% smoking cessation rate at 12 months, when patients participated in a nursing inpatient smoking cessation program with telephone follow-up (p < .01) compared with a group of patients hospitalized for an acute myocardial infarction (AMI; 32% smoking cessation). Dornelas, Sampson, Gray, Waters, and Thompson (2000) found a 67% smoking cessation rate at 6 months, compared with 43% in the usual care group, when AMI patients participated in a psychologist-led inpatient intervention with telephone follow-up, based on the stages of change of the Transtheoretical Model (p = .05). This latter intervention was, however, evaluated only in persons having had an AMI. Also, self-reported smoking abstinence was not confirmed by biochemical analyses, and the contribution of the telephone follow-up was not evaluated.

Within an inpatient environment, nurses are the health care professionals who have the most contact with patients. As smoking is prohibited on the wards, hospital

Maud-Christine Chouinard, PhD, RN, is an Assistant Professor, Module des sciences infirmières et de la santé, Université du Québec à Chicoutimi, Québec, Canada.
Sylvie Robichaud-Ekstrand, PhD, RN, is an Associate Professor, Faculty of Nursing, Université de Montréal, Quebec, Canada.
environments become ideal settings for individuals to stop smoking and to gain the appropriate skills needed to remain nonsmokers after discharge (Cote, 2000). It is anticipated that once individuals have acquired initial smoking relapse-prevention skills (through participation in an inpatient smoking cessation program), the telephone follow-up by the nurse along with family support may be all that is required to prevent relapse. The readiness to stop smoking, and the level at which the intervention is adapted to participants’ characteristics, may contribute to the effectiveness of the intervention (Andersen & Keller, 2002). Introducing such a tailored program may reach a considerable number of patients with CVD within a hospital environment, and become more effective after discharge (Johnson, Budz, Mackay, & Miller, 1999). Also, increased participation rates and improved effectiveness would create a greater impact on smoking cessation at the population level (Abrams et al., 1996). However, several studies led by nurses did not yield the anticipated results (Bolman, de Vries, & van Breukelen, 2002; Hajek, Taylor, & Mills, 2002; Reid et al., 2003). The lack of tailoring smoking cessation programs, the short duration span, and the limited sample size are possible contributing factors.

The controversial results in nurse-led smoking cessation programs and certain methodological issues not considered warrant further investigations (Sebregts, Falger, & Bär, 2000; Wiggers, Smets, de Haes, Peters, & Legemate, 2003). The purpose of this study was to evaluate the effectiveness of a nursing inpatient smoking cessation program with telephone follow-up, according to the Transtheoretical Model. In addition, this study was designed to delineate the effects of the added value of the telephone follow-up.

**Theoretical Framework**

The study was conceptualized using the Transtheoretical Model (Prochaska & DiClemente, 1983), and the smoking cessation intervention was developed by integrating elements of this model. Because multiple attempts may be necessary to become a nonsmoker, smoking cessation was considered a process rather than an end product (smoker vs. nonsmoker). It is believed that progress to higher levels of change required particular cognitive and behavioral processes of change. The utilization of specific processes of change reflects the individual’s readiness to become a nonsmoker. Stage of change is a reflection of this readiness, and can be used to evaluate the effectiveness of a smoking cessation program (Rustin & Tate, 1993). Therefore, the proposed smoking cessation nursing intervention was considered effective if a smoker progressed to further stages of change or quit smoking.

Given that most people do not follow a linear path when modifying a behavior, the Transtheoretical Model is schematically represented as a spiral, which indicates that many attempts may be necessary before modifying and maintaining a healthy behavior. Individuals can regress to a previous stage, learn from it, and eventually reach the maintenance stage. Five stages of change are identified in the Transtheoretical Model (Prochaska, Redding, & Evers, 2002). In the precontemplation stage, there is no intention to modify behavior within the next 6 months. Contemplation signifies that individuals are becoming aware that smoking is problematic, but still feel ambivalent about quitting. They wish to quit smoking either within the next 6 months or within the next month (if they state never having attempted to quit smoking for at least 24 hr over the past year). Preparation is the stage that combines intention and behavioral efforts. Individuals at this stage initiate actions to stop smoking in the next month and back up their intention by having attempted to quit during the past year. Action is the 6-month time frame where individuals alter their environment, experiences, and relationships, and stop smoking. However, they are at high risk of relapsing because habits are not yet established. Maintenance is the 6-month and over time frame where an individual remains a nonsmoker. It requires motivation and skills to prevent relapse.

A number of scientific articles aimed at health care professionals underline the importance of integrating the information generated by the Transtheoretical Model into smoking cessation programs (Cole, 2001; Cote, 2000; Coward, 1999). Yet, the effects of smoking cessation programs based on this model remain controversial (Andersen, Keller, & McGowan, 1999; Riemsma et al., 2003; Spencer, Pagell, Hallion, & Adams, 2002).

In this study, smoking cessation at 6 months was one desired outcome. More subtle cognitive changes that occur during the normal process of quitting smoking may occur as well (Abrams, 1993; Velicer, Rossi, DiClemente, & Prochaska, 1996). Consequently, progress to further stages of change was chosen as a secondary outcome.

The objective of the study was to test the following hypotheses: Smokers with CVD receiving a nurse-delivered inpatient smoking cessation counseling intervention (based on stages of change) with telephone follow-up (Group 1) or without (Group 2) will present higher point-prevalent smoking abstinence, higher rates of continuous abstinence from smoking, and a better progression to ulterior stages of change at 2 and 6 months after hospital discharge than the usual care group (Group 3). The inpatient smoking cessation counseling intervention with telephone follow-up (Group 1) was expected to produce significantly better results than the inpatient smoking cessation counseling intervention only (Group 2). Point-prevalent smoking abstinence referred to not having smoked for the past 7 days, and obtaining a negative biochemical result at the time of measurement. Continuous smoking abstinence indicated having refrained from smoking since hospital discharge. It was anticipated that the initial stage of change would predict smoking cessation.

**Method**

**Sample**

**Selection Criteria** The following inclusion criteria were required for participation in the study: (a) adult (18 years and older); (b) hospitalized for a CVD (myocardial infarction, angina, heart failure, or peripheral vascular disease); (c) smoker (having smoked at least one cigarette in the past
month); (d) the ability to communicate in French; (e) local resident; (f) a telephone available at home; (g) plan of hospital discharge to home; (h) no mental or physical disabilities that would impede participation.

**Sample Size** Sample size was determined using the Pass 2000 software. For a chi-square Pearson test, the unilateral proportion differential (β) was set at 25% for smoking cessation and progress through stages of change between the groups receiving the inpatient counseling with telephone follow-up and the one receiving usual care. With the critical alpha value set at 5% (type I error), and a power (1 - β) of 80% (type II error), the sample size was established at 52 participants per group.

**Procedure** The research protocol and consent form were approved by scientific and ethical committees. At patient admission, medical records were reviewed, and potential participants were identified by verifying medical diagnosis and smoking status. The study was explained to all eligible participants whose condition was stable. If family members were present, the study was explained to them as well. All patients who wished to participate in the study signed a consent form. Data collected during the inpatient period included the following: (a) stages of change, (b) data on nicotine dependence, (c) clinical and sociodemographic data, and (d) level of psychological distress. For comparison purposes (participants vs. nonparticipants), some patients agreed to provide basic clinical and sociodemographic information, and they signed the respective consent form.

A convenience sample of participants was created. To prevent contamination between groups, cluster randomization was used (Haucik, Gilliss, Donner, & Gortner, 1991). This was done by first randomly assigning individuals to predetermined clusters of three to six subjects. The group assignment (inpatient counseling with telephone follow-up, inpatient counseling only, usual care) was then randomly assigned to each of these clusters. Individuals not familiar with the study were in charge of the randomization procedure, which included inserting the information into envelopes that were sealed and would be opened by the investigator only at the time of recruitment.

The experimental groups (inpatient counseling with telephone follow-up; inpatient counseling only) met the two intervention nurses. The duration of each session varied according to the individual’s health status and specific needs. The usual care group received general advice on smoking cessation. After discharge, individuals in the inpatient counseling with telephone follow-up group received a total of six phone calls over the next 2 months in the following sequence: two calls during the first week and one call per week over the next 2 weeks, followed by one call every 2 weeks.

A nursing research assistant met each participant at his or her home 2 and 6 months after hospital discharge. Participants completed the same instruments and questionnaires as at baseline. Participants who declared not having smoked over the past week underwent a biochemical validation test.

**Smoking Cessation Intervention** The following principles pertaining to the smoking cessation intervention were applied to the experimental groups: (a) the smoking cessation intervention was adapted to each individual’s health condition and needs; (b) a significant family member, preferably living with the participant, was involved; (c) an explanation of the stages of change according to the Transtheoretical Model was presented to the participants and to a significant family member; (d) information related to how family members could concretely provide support to the CVD patients was provided (Cohen, 1992); and (e) the importance of remaining a non-smoker was emphasized by the nurse (Agency for Health Care Policy and Research, 1996).

The inpatient counseling and telephone follow-up was based on the Transtheoretical Model (Prochaska, DiClemente, & Norcross, 1997). During the precontemplation stage, the smoking cessation intervention focused on increasing the perceived cons of smoking (within the decisional balance), and encouraged the use of experiential processes of change (cognitive and affective processes). At the contemplation stage, the intervention aimed at increasing the cons of smoking and diminishing the pros. Experimental, as well as behavioral, processes of change were incorporated (actions undertaken by the smoker). At the preparation stage, four principles were targeted: (a) reversal of the decisional balance (more cons than pros of smoking); (b) the more extensive use of behavioral processes of change; (c) the reinforcement of self-efficacy; and (d) the acquisition of relapse-prevention skills. The action stage focused on increasing the use of behavioral processes, reinforcing self-efficacy, providing relapse-prevention skills, and reinforcing smoking cessation attempts by congratulating the individual for deciding to stop smoking and to remain a nonsmoker. During each telephone interview, the intervention was tailored to the actual stage of change and smoking habits.

The inpatient counseling session lasted 40 min on the average (SD = 8.1), ranging from 10 to 60 min. Family members, spouses in particular (n = 44, 84.6%), accompanied the patients in 46.4% (n = 52) of the interviews. Seventy-five percent of the 56 participants in the inpatient cessation program with telephone follow-up received all six telephone calls. In the other cases, three preferred to interrupt the calls, but desired to continue with the study. As for the other cases, fewer calls were made because the participants were difficult to reach.

Patients who were physically dependent on nicotine, and their physical condition allowed it, were oriented toward pharmacological aids. Therefore, 29 participants used nicotine patches, 6 chewed nicotine gum, and 6 were prescribed bupropion (Zyban®). Significantly more participants in the inpatient counseling with telephone follow-up group used pharmacological aids (n = 22, 39.2%), than the inpatient counseling only (n = 13, 23.2%), and usual care (n = 6, 10.7%) groups (p < .01). However, among these, only nine patients (22%) completely followed the 3-month pharmacological treatment. This may indicate that, when followed by the nurse, patients with CVD are better informed about nicotine cessation drug adjuncts. Despite the fact that only 17% (n = 29) of the participants...
used nicotine cessation drug adjuncts, cost was not an issue (i.e., these pharmacological aids are covered by a universal medical insurance plan). The main reticence from doctors and patients for using nicotine cessation drug adjuncts was safety and their effectiveness in smoking cessation.

**Instruments**

Instruments used to gather information related to smoking included Smoking Stages of Change—short form (administered at baseline and at 2 and 6 months); the Fagerström Test for Nicotine Dependence and l’Indice de détresse psychologique de l’Enquête Santé Québec (IDPESQ; Quebec Health Survey Psychological Distress Index; completed at baseline). Smoking status was evaluated by self-reports and, if participants declared themselves as being nonsmokers, this was confirmed using biochemical tests. Additional clinical and sociodemographic questionnaires were administered at baseline and at 2 and 6 months.

**Stages of Change** Stage of change was identified using a French version of the Smoking: Stage of Change (short form). This instrument is composed of four questions, which evaluate the person’s intention to quit smoking within a specific time interval, hence identifying the specific stage of change according to the Transtheoretical Model (Figure 1). Because this questionnaire is not a psychological evaluation, no psychometric properties are reported in the literature. Nevertheless, stages of change have predicted smoking cessation or relapse (DiClemente & Prochaska, 1985). When analyzing movement to different stages of change, a score of 1 was given to patients who progressed to higher stages, while those who remained at the same stage, or relapsed to a previous stage, were given a score of 0.

**Nicotine Dependence** Nicotine dependence was measured using the French version of the Fagerström Test for Nicotine Dependence (Heatherton, Kozlowski, Frecker, & Fagerström, 1991). The scale comprises six questions. Its sum reflects the severity of physical dependence on nicotine. Scores ranging from 7 to 10 indicate strong nicotine dependence; scores of 4–6 signify a moderate dependence, while those below 4 reflect little dependence. With a Cronbach alpha coefficient of .61, the scale displays moderate internal reliability. This instrument is used extensively in the clinical practice. In this study, its internal reliability was .59.

**Psychological Distress** The severity of the most prevalent symptoms in depression or anxiety was measured by this 14-item instrument. A provincewide survey in Quebec (Canada) provided sufficient data to establish the psychometric properties of this instrument with regard to construct validity, criterion-related validity, and reliability (Préville, Boyer, Potvin, Perrault, & Legare, 1992). The mean Cronbach alpha coefficient was .89. In this study, Cronbach alpha coefficient was .90.

**Smoking Status** Self-reports on smoking cessation were validated using a Home Health Testing® urinary continence test. This immuno-test detects active smoking elements in the urine if a person has smoked during the previous 4 days. It has a 5% error rate. In the event that some participants conjunctly used a nicotine substitute, smoking status was further validated with a Bedfont carbon monoxide exhalation test, Smokerlyser®. In an earlier study, biochemical measures enhanced the validity of the self-reports, especially if the individual was at a high risk of relapsing (Patrick et al., 1994). It is likely that this study represented a similar situation.

---

**FIGURE 1.** Interpretation algorithm of the smoking behavior stages of change scale.
Data Analysis

Chi-square analyses and analyses of variance (ANOVAs) were utilized to determine if randomization provided equivalent groups (i.e., the participant’s sociodemographic and clinical characteristics at baseline) were not significantly different. Similar analyses were used to compare patients who consented to participate in the study, and those who declined but agreed to provide some information.

Smoking cessation rates were calculated using all available follow-up data. It was recorded if participants were lost to follow-up, or discontinued their participation when they restarted smoking. No cluster analyses were performed because of the small number of participants per cohort (3–6 participants). To prevent contamination between groups (i.e., to avoid patients in different groups benefiting from similar smoking cessation information), a probabilistic sampling method was utilized. Contingency tables with $2 \times 3$ chi-square analyses were used for discrete variables such as point-prevalent smoking abstinence, continuous smoking abstinence, and stage progression at 2 and 6 months. If a significant difference was found between the three groups (inpatient counseling with telephone follow-up, inpatient counseling, usual care), $2 \times 2$ chi-square analyses were performed.

### TABLE 1. Characteristics of Participants Versus Nonparticipants ($N = 225$)

<table>
<thead>
<tr>
<th></th>
<th>Participants ($n = 168$)</th>
<th>Nonparticipants ($n = 57$)</th>
<th>$p^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>45 (26.8)</td>
<td>20 (35.1)</td>
<td>.23</td>
</tr>
<tr>
<td>Men</td>
<td>123 (73.2)</td>
<td>37 (64.9)</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>124 (73.8)</td>
<td>38 (66.7)</td>
<td>.30</td>
</tr>
<tr>
<td>Education (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–12</td>
<td>131 (78)</td>
<td>53 (93)</td>
<td>.01</td>
</tr>
<tr>
<td>13 and more</td>
<td>37 (22)</td>
<td>4 (7)</td>
<td></td>
</tr>
<tr>
<td>Work status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>60 (35.7)</td>
<td>12 (21.1)</td>
<td>.12</td>
</tr>
<tr>
<td>Retired</td>
<td>65 (38.7)</td>
<td>28 (49.1)</td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>43 (25.6)</td>
<td>17 (29.8)</td>
<td></td>
</tr>
<tr>
<td>Family income$^b$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below $14,999$</td>
<td>29 (18.1)</td>
<td>19 (34.5)</td>
<td>.01</td>
</tr>
<tr>
<td>$15,000–$59,999</td>
<td>104 (65.0)</td>
<td>33 (60)</td>
<td></td>
</tr>
<tr>
<td>Above $60,000$</td>
<td>27 (16.9)</td>
<td>3 (5.5)</td>
<td></td>
</tr>
<tr>
<td>Principal diagnosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>40 (23.8)</td>
<td>11 (19.3)</td>
<td>.37</td>
</tr>
<tr>
<td>Angina</td>
<td>61 (36.3)</td>
<td>26 (45.6)</td>
<td></td>
</tr>
<tr>
<td>Heart failure</td>
<td>11 (6.5)</td>
<td>1 (1.8)</td>
<td></td>
</tr>
<tr>
<td>Peripheral vascular disease</td>
<td>56 (33.3)</td>
<td>19 (33.3)</td>
<td></td>
</tr>
<tr>
<td>Previous quit attempt</td>
<td>141 (83.9)</td>
<td>44 (77.2)</td>
<td>.25</td>
</tr>
<tr>
<td>Medical advice</td>
<td>124 (73.8)</td>
<td>38 (66.7)</td>
<td>.30</td>
</tr>
<tr>
<td>Stage of change</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precontemplation</td>
<td>20 (11.9)</td>
<td>18 (31.6)</td>
<td>.01</td>
</tr>
<tr>
<td>Contemplation</td>
<td>18 (10.7)</td>
<td>4 (7)</td>
<td></td>
</tr>
<tr>
<td>Preparation</td>
<td>121 (72)</td>
<td>30 (52.6)</td>
<td></td>
</tr>
<tr>
<td>Action</td>
<td>9 (5.4)</td>
<td>5 (8.8)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>55.9 (10.4)</td>
<td>60.0 (11.7)</td>
<td>.01</td>
</tr>
<tr>
<td>Length of stay (days)</td>
<td>7.15 (11.2)</td>
<td>7.72 (10.4)</td>
<td>.74</td>
</tr>
<tr>
<td>Age at smoking initiation (years)</td>
<td>15.21 (4.5)</td>
<td>14.14 (5.0)</td>
<td>.05</td>
</tr>
</tbody>
</table>

$^a$Group comparisons t test for continuous and $\chi^2$ for nominal and categorical variables.

$^b$The total number of participants is not always equal to 225 because of refusals to answer.
Multiple logistic regression analyses identified the variables that best predicted smoking abstinence and progression to ulterior stages of change. All 21 variables at baseline were entered, including the duration of the inpatient smoking cessation intervention and family member participation (Forward method). Thereafter, significant variables were incorporated into the regression analyses (Enter method).

Results

Participants
The convenience sample was composed of 168 patients with CVD recruited from a cardiology unit, within a regional tertiary hospital in the province of Quebec (Canada). Between December 2001 and August 2002, 1,354 consecutive patients hospitalized for CVD were screened; 267 (19.7%) were identified as smokers (Figure 2).

Data were obtained from persons who declined to fully participate in the study. This information was compared to that of participants (Table 1), and indicated that those who declined to participate were older, less educated, had lower family incomes, and had started smoking at a younger age. Of these participants, more were found at the precontemplation stage (i.e., they did not envision quitting smoking in the next 6 months). Comparative analyses at baseline indicated that participants in all three experimental groups presented similar sociodemographic and clinical characteristics (Table 2), as well as similar smoking characteristics (Table 3).

Validation
At 2 months, 61 individuals stated they were nonsmokers. Biochemical confirmation was obtained in 55 (3 were not tested for reasons related to the investigators, and 3 others

| TABLE 2. Sociodemographic and Clinical Characteristics of Sample at Baseline |

<table>
<thead>
<tr>
<th>Inpatient Counseling With Telephone Follow-Up (n = 56)</th>
<th>Inpatient Counseling Only (n = 56)</th>
<th>Usual Care (n = 56)</th>
<th>Total (N = 168)</th>
<th>p^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>13 (23.2)</td>
<td>11 (19.6)</td>
<td>21 (37.5)</td>
<td>45 (26.8)</td>
</tr>
<tr>
<td>Men</td>
<td>43 (76.8)</td>
<td>45 (80.4)</td>
<td>35 (62.5)</td>
<td>123 (73.2)</td>
</tr>
<tr>
<td>Married</td>
<td>40 (71.4)</td>
<td>44 (78.6)</td>
<td>40 (71.4)</td>
<td>124 (73.8)</td>
</tr>
<tr>
<td>Education (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–12</td>
<td>45 (80.4)</td>
<td>41 (73.2)</td>
<td>45 (80.4)</td>
<td>131 (78.0)</td>
</tr>
<tr>
<td>13 and more</td>
<td>11 (19.6)</td>
<td>15 (26.8)</td>
<td>11 (19.6)</td>
<td>37 (22.0)</td>
</tr>
<tr>
<td>Work status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>23 (41.1)</td>
<td>24 (42.9)</td>
<td>13 (23.2)</td>
<td>60 (35.7)</td>
</tr>
<tr>
<td>Retired</td>
<td>18 (32.1)</td>
<td>22 (39.3)</td>
<td>25 (44.6)</td>
<td>65 (38.7)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>15 (26.8)</td>
<td>10 (17.9)</td>
<td>18 (32.1)</td>
<td>43 (25.6)</td>
</tr>
<tr>
<td>Family income</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below $14,999</td>
<td>10 (18.2)</td>
<td>6 (11.1)</td>
<td>13 (25.5)</td>
<td>29 (18.1)</td>
</tr>
<tr>
<td>$15,000–$59,999</td>
<td>32 (58.2)</td>
<td>38 (70.4)</td>
<td>34 (66.7)</td>
<td>104 (65.0)</td>
</tr>
<tr>
<td>Above $60,000</td>
<td>13 (23.6)</td>
<td>10 (18.5)</td>
<td>4 (7.8)</td>
<td>27 (16.9)</td>
</tr>
<tr>
<td>Principal diagnosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myocardial infarction</td>
<td>16 (28.6)</td>
<td>12 (21.4)</td>
<td>12 (21.4)</td>
<td>40 (23.8)</td>
</tr>
<tr>
<td>Angina</td>
<td>22 (39.3)</td>
<td>22 (39.3)</td>
<td>17 (30.4)</td>
<td>61 (36.3)</td>
</tr>
<tr>
<td>Heart failure</td>
<td>4 (7.1)</td>
<td>3 (5.4)</td>
<td>4 (7.1)</td>
<td>11 (6.5)</td>
</tr>
<tr>
<td>Peripheral vascular disease</td>
<td>14 (25.0)</td>
<td>19 (33.9)</td>
<td>23 (41.1)</td>
<td>56 (33.3)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>M (SD)</th>
<th>Age</th>
<th>Length of stay (days)</th>
<th>Psychological distress</th>
</tr>
</thead>
<tbody>
<tr>
<td>M (SD)</td>
<td>54.0 (10.8)</td>
<td>56.5 (10.3)</td>
<td>57.25 (9.9)</td>
</tr>
<tr>
<td>Length of stay (days)</td>
<td>7.6 (13.1)</td>
<td>7.5 (13.0)</td>
<td>6.4 (6.5)</td>
</tr>
<tr>
<td>Psychological distress</td>
<td>1.50 (0.48)</td>
<td>1.46 (0.45)</td>
<td>1.54 (0.48)</td>
</tr>
</tbody>
</table>

^aGroup comparisons ANOVA for continuous and \( \chi^2 \) for categorical variables.
refused to undergo testing and were considered smokers). Of the 55 participants tested, 38 (69.1%) underwent urine tests and presented negative results (confirmed by a negative carbon monoxide test in 51.5%, \( n = 20 \)), and 14 (25.5%) underwent carbon monoxide tests and showed negative results. Finally, 3 persons (5.5%) showed positive results and were reclassified as smokers. At 6 months, 55 individuals declared being nonsmokers; however, 6 of them refused to undergo testing, and were classified as smokers. Of the 49 remaining individuals, 42 (85.7%) had negative results after urine test (and 13 [31%] were confirmed by carbon monoxide testing), 5 (10.2%) underwent carbon monoxide testing with negative results, and 2 individuals were not tested.

### Point-Prevalent Smoking Abstinence

At 2 months, individuals in the inpatient counseling with telephone follow-up group significantly abstained from smoking compared with the inpatient counseling and the usual care groups: 44.4% (\( n = 24 \)) compared with 33.3% and 23.6%, respectively, \( \chi^2(2, N = 163) = 5.28; p < .07 \). At 6 months, the difference was slightly higher: 41.5% (\( n = 22 \)), 30.2% (\( n = 16 \)), and 20.0% (\( n = 11 \)) point-prevalent smoking abstinence, in the respective groups, \( \chi^2(2, N = 161) = 5.90; p < .05 \) (Table 4).

### Continuous Smoking Abstinence

At 2 months, 42.6% (\( n = 23 \)) of the inpatient counseling with telephone follow-up group quit smoking and
remained nonsmokers, compared with 29.6% (n = 16) in the inpatient counseling group, and 21.8% (n = 12) in the usual care group (Table 4). Although this difference appears clinically significant, it is not statistically significant, \( \chi^2(2, N = 163) = 5.57; p < .06 \). At 6 months, the continuous abstinence rate was twice as high in the intervention groups as in the usual care group (24.5%, 24.5%, and 12.7%, respectively), but this difference was still not significant, \( \chi^2(2, N = 161) = 3.01; p = .21 \).

**Progression to Ulterior Stages of Change**

At 2 months, 48.1% (n = 26) of the inpatient counseling with telephone follow-up group progressed to ulterior stages of change, compared to 29.6% (n = 16) in the inpatient counseling group, and 27.3% (n = 15) in the usual care group, \( \chi^2(2, N = 163) = 6.23; p < .04 \). At 6 months, these differences were greater, and there existed a smaller number in the usual care group having progressed to ulterior stages of change, \( \chi^2(2, N = 161) = 8.05; p < .02 \) (Table 4).

Further data analyses indicate that no precontemplator or contemplator receiving usual care progressed to subsequent stages of change at 6 months, while 25% (n = 3) in the inpatient intervention groups did progress, \( \chi^2(2, N = 36) = 6.36; p = .04 \). While no significant difference in the levels of distribution in the stages of change was apparent between groups at 2 months, \( \chi^2(6, N = 133) = 9.71; p = .07 \).

### TABLE 3. Smoking Characteristics of Sample at Baseline

<table>
<thead>
<tr>
<th></th>
<th>Inpatient Counseling With Telephone Follow-Up (n = 56)</th>
<th>Inpatient Counseling Only (n = 56)</th>
<th>Usual Care (n = 56)</th>
<th>Total (N = 168)</th>
<th>p^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previous quit attempt</td>
<td>47 (83.9)</td>
<td>47 (83.9)</td>
<td>47 (83.9)</td>
<td>141 (83.9)</td>
<td>1.00</td>
</tr>
<tr>
<td>Medical advice about smoking cessation</td>
<td>47 (83.9)</td>
<td>38 (67.9)</td>
<td>39 (68.6)</td>
<td>124 (73.8)</td>
<td>.11</td>
</tr>
<tr>
<td>Stage of change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Precontemplation</td>
<td>6 (10.7)</td>
<td>5 (8.9)</td>
<td>9 (16.1)</td>
<td>20 (11.9)</td>
<td>.86</td>
</tr>
<tr>
<td>Contemplation</td>
<td>5 (8.9)</td>
<td>7 (12.5)</td>
<td>6 (10.7)</td>
<td>18 (10.7)</td>
<td></td>
</tr>
<tr>
<td>Preparation</td>
<td>43 (76.8)</td>
<td>40 (71.4)</td>
<td>38 (67.9)</td>
<td>121 (72.0)</td>
<td></td>
</tr>
<tr>
<td>Action</td>
<td>2 (3.6)</td>
<td>4 (7.1)</td>
<td>3 (5.4)</td>
<td>9 (5.4)</td>
<td></td>
</tr>
</tbody>
</table>

\[ M (SD) \]

<table>
<thead>
<tr>
<th>Age at smoking initiation</th>
<th>15.8 (5.9)</th>
<th>15.6 (3.6)</th>
<th>15.1 (3.6)</th>
<th>15.5 (4.5)</th>
<th>.74</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addiction</td>
<td>5.6 (2.3)</td>
<td>4.5 (2.8)</td>
<td>5.1 (2.1)</td>
<td>5.1 (2.5)</td>
<td>.07</td>
</tr>
</tbody>
</table>

^aGroup comparisons ANOVA for continuous and \( \chi^2 \) for categorical and nominal variables.

### TABLE 4. Smoking Cessation Results at 2 and 6 Months

<table>
<thead>
<tr>
<th></th>
<th>Inpatient Counseling With Telephone Follow-Up</th>
<th>Inpatient Counseling Only</th>
<th>Usual Care</th>
<th>( \chi^2 )</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point-prevalent smoking abstinence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 months</td>
<td>24 (44.4)</td>
<td>18 (33.3)</td>
<td>13 (23.6)</td>
<td>5.28</td>
<td>.07</td>
</tr>
<tr>
<td>6 months</td>
<td>22 (41.5)</td>
<td>16 (30.2)</td>
<td>11 (20.0)</td>
<td>5.90</td>
<td>.05</td>
</tr>
<tr>
<td>Continuous smoking abstinence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 months</td>
<td>23 (42.6)</td>
<td>16 (29.6)</td>
<td>12 (21.8)</td>
<td>5.57</td>
<td>.06</td>
</tr>
<tr>
<td>6 months</td>
<td>13 (24.5)</td>
<td>13 (24.5)</td>
<td>7 (12.7)</td>
<td>3.01</td>
<td>.21</td>
</tr>
<tr>
<td>Progress through stages of change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 months</td>
<td>26 (48.1)</td>
<td>16 (29.6)</td>
<td>15 (27.3)</td>
<td>6.23</td>
<td>.04</td>
</tr>
<tr>
<td>6 months</td>
<td>s23 (43.4)</td>
<td>17 (32.1)</td>
<td>10 (18.2)</td>
<td>8.05</td>
<td>.02</td>
</tr>
</tbody>
</table>
Abstract

This study evaluated the efficacy of a nurse-led inpatient smoking cessation program with telephone follow-up for patients with coronary artery disease (CAD) and other cardiovascular diseases (CVD). The program consisted of individualized counseling, group classes, and telephone follow-up for 6 months. The primary outcome measure was point-prevalent smoking abstinence at 6 months.

Results

A total of 332 participants were recruited from four hospitals, and 271 participants completed the study. The smoking abstinence rate at 6 months was 41.5% for the inpatient counseling with telephone follow-up group, compared to 29.6% for the inpatient counseling only group and 23.1% for the usual care group. Multivariate logistic regression analyses revealed that the initial stage of readiness to change and group assignment were significant predictors of smoking abstinence at 6 months. The initial stage of readiness to change was the only variable that predicted point-prevalent smoking abstinence at 6 months.

Discussion

The study findings suggest that individualized counseling with telephone follow-up is an effective strategy for smoking cessation in patients with CVD. The program may be particularly beneficial for patients with CAD and other CVD, as it can be tailored to meet the specific needs of each patient. Future research should focus on evaluating the long-term effects of the program and identifying additional strategies to improve smoking cessation outcomes.

Graphical abstract

The graphical abstract illustrates the stage of change by group, at baseline, 2 and 6 months posthospital discharge.
studied the effects of such a smoking cessation program in individuals with peripheral vascular disease (Clarke & Aish, 2002; Galvin, Webb, & Hillier, 2001; Power, Brown, & Makin, 1992).

In most smoking cessation studies highly motivated individuals were recruited, but participants were included in this study regardless of their level of motivation or readiness to quit smoking. This might explain the relatively lower smoking abstinence rates (Pierce, Farkas, & Gilpin, 1998). It is important to note that participants who presented lower levels of readiness to quit smoking at baseline progressed more readily to further stages of change. In this study, all participants in the inpatient counseling with telephone follow-up group were at the precontemplation stage at baseline and progressed to ulterior stages of change at 2 or 6 months. While some individuals were still smoking, a few were seriously thinking of quitting within the next 6 months; others were preparing their environment or seeking support from others to help them quit. According to the stages of change made, as participants progressed to ulterior stages of change, their chances of quitting smoking at 6 months after hospital discharge became higher (Prochaska et al., 2002). This was confirmed as stage of change at baseline was predictive of point-prevalent smoking abstinence at 6 months. Similar to the results of the current study, Taylor et al. (1990) found that high intention to quit smoking at hospital discharge predicted smoking cessation at 12 months.

Participation in this study was influenced by the stage of change because 28% of the hospitalized patients who declined to enroll in the study were predominantly at the precontemplation stage. This result was observed also by Clarke and Aish (2002).

It was reported earlier that significant increases in smoking abstinence resulted when inpatient smoking cessation programs included a telephone follow-up (Dornelas et al., 2000; Taylor et al., 1990). In this study, the telephone follow-up was beneficial, particularly for participants who had restarted to smoke after hospital discharge (point-prevalent smoking abstinence vs. continuous abstinence). Thus, telephone follow-up permitted a better progression to ulterior stages of change, which ultimately led to higher smoking cessation rates. It doubled the point-prevalent smoking abstinence rate (21.5% above usual care and 11.3% above inpatient counseling only). These results are supported by Hajek et al. (2002), who obtained no significant results with a short inpatient smoking intervention in individuals with heart disease (MI and angina). This pattern suggests that smokers normally benefit from continuous support from health care professionals after hospital discharge, as smokers go through several attempts before becoming permanent nonsmokers. It also reinforces the logic for representing the Transtheoretical Model as a spiral that refers to the multiple attempts undertaken in behavior modification.

In this study, nurses encouraged family members to participate in the inpatient smoking cessation counseling sessions; however, few family members got involved—the main reason being the short hospitalization stay. Further analyses indicated that family member participation was not a predictor of smoking abstinence or progression to ulterior stages of change. A more intensive and sustained smoking cessation intervention may have influenced social support dispensed by family members. In addition, duration of the inpatient smoking cessation counseling and telephone follow-up was not a predictor. This further suggests that a relatively brief smoking cessation intervention could be as effective.

The results of this study can be generalized only to the French-speaking Canadian population with CVD. Furthermore, as three quarters of the sample were men, the results cannot be generalized to women. Also, despite the effectiveness of the randomized method, contamination between groups could have occurred—for example, during

<table>
<thead>
<tr>
<th>Variables</th>
<th>Stage of change at baseline</th>
<th>OR</th>
<th>95% CI</th>
<th>Progression in Stages of Change at 6 Months</th>
<th>OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precontemplation</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contemplation</td>
<td>.58</td>
<td>0.05–6.46</td>
<td>2.13</td>
<td>0.25–18.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preparation</td>
<td>2.27</td>
<td>0.61–8.40</td>
<td>3.87</td>
<td>0.84–17.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Action</td>
<td>11.16*</td>
<td>1.69–73.68</td>
<td>11.07*</td>
<td>1.48–82.68</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. OR = odds ratio. CI = confidence intervals.

*p < .05. **p < .01.
a rehospitalization period, or participation in a cardiac rehabilitation program. No restriction existed for patients seeking information on smoking cessation once they were discharged from the hospital.

The Transtheoretical Model was useful in tailoring the smoking cessation program according to selected cognitive and behavioral components. Results in this study indicate that this nursing inpatient smoking cessation program, which was tailored to patients’ readiness to change, increases point-prevalent smoking abstinence at 6 months. In addition, patients with CVD who are still smoking at 6 months after hospital discharge demonstrate higher levels of motivation to quit smoking in the near future.

References


A Meta-Analysis of the Effect of Hospital-Based Case Management on Hospital Length-of-Stay and Readmission

Young-Ju Kim ▼ Karen L. Soeken

Background: Although many hospital-based case management (CM) interventions have been studied, there is little work summarizing the effectiveness of these studies.

Objectives: The purpose of this study was to investigate the effect of hospital-based CM compared with usual care on length of hospital stay and readmission rate.

Method: A meta-analytic method was employed to analyze the effect sizes of CM intervention on outcomes. Eligible studies were retrieved using computerized database searches, footnote chasing, and contact with content experts. The authors reviewed the final 12 studies, and the effect size, 95% confidence interval (CI), sensitivity, homogeneity, and publication bias were analyzed.

Results: The overall average weighted effect size on length of stay (LOS) was 0.094 with a 95% CI of –0.032 to 0.220. The overall odds ratio for readmission was 0.87 with a 95% CI of 0.69 to 1.04. Overall, hospital-based CM interventions were not significantly effective in reducing LOS and readmissions. However, CM for patients with heart failure (effect size of 0.241 with a 95% CI of 0.012 to 0.470) was significantly effective in reducing LOS, although it was not effective for stroke patients (effect size of –0.226 with a 95% CI of –0.542 to 0.089) and frail elders (effect size of 0.126 with a 95% CI of –0.073 to 0.324). Analysis indicated that in this meta-analysis publication bias was unlikely.

Discussion: The findings of this meta-analysis demonstrate a 6% decrease in readmission rate for patients who received hospital-based CM interventions. Further meta-analytic studies are needed to investigate the effectiveness of CM on other outcomes.

Key Words: case management • length of stay • meta-analysis • readmission

Over the past 15 years, models of case management (CM) have been developed in healthcare to maintain or improve the quality of care in a milieu of financial constraints in the United States. These models often are categorized by the healthcare setting as follows: hospital-based, hospital-to-community–based, and community-based models (Lamb, 1992). The focus in this study is on hospital-based models that are designed to improve care delivery outcomes by ensuring that patients receive what they need when they need it (Cook, 1998).

Wide variations exist in the components of hospital-based CM. However, in many discussions of CM, six core components emerged: assessment, education, collaboration, discharge planning, linkage, and monitoring (Eckett, Vassallo, & Flett, 1996; Kanter, 1989; Rothman, 1991). Case management requires a comprehensive assessment of the patient and family’s social, physical, and psychological functioning (Rothman). Without accurate assessment, CM interventions frequently are unsuccessful (Kanter). Patients themselves have reported multiple concerns regarding their progress, activity, medication, and pain control (Boyle, Nance, & Passau-Buck, 1992). They require quality information that will prevent future disease-related episodes. Therefore, patient and family education during hospitalization is a large part of the nurse case manager role.

Case management involves collaborative multidisciplinary practice. Nurse case managers are required to plan and coordinate healthcare services that respond to the individualized needs of patients and families (Gibson, Martin, Johnson, & Miller, 1994). Early discharge planning, defined as beginning within 48 hours of admission to the hospital (Dugan & Mosel, 1992), is associated with the process of assessing the unmet needs of patients and developing a coordinated care plan (Feather, 1993).

Linking patients with needed resources is common among case managers. The formal linkage activities include clarifying the service needs, making telephone contact, orienting the patient, and so forth (Rothman, 1991). The purpose of monitoring is to assess the suitability of provisions made to sustain patients in the place where they

Young-Ju Kim, MSN, RN, is Doctoral Candidate; and Karen L. Soeken, PhD, is Professor, School of Nursing, University of Maryland, Baltimore.
are discharged (Rothman, 1991). It can be done through telephoning, visiting, or having the patient phone the case manager. The specific components of CM that are implemented may vary by hospital. Coordination of all core components on a continuum is the key for successful implementation of CM.

There are a number of care delivery outcomes that are expected from implementing hospital-based CM. Most of the models commonly purport to enhance quality of care, increase patient satisfaction, reduce length of hospital stay, and reduce costs of care (Cook, 1998; Renholm, Leinon-Kilpi, & Suominen, 2002). Cost containment and cost reduction are key components in the managed care environment (Parr, 1996). In addition, LOS and resource consumption became a top priority for hospitals with legislation creating the prospective payment system (Fields, 1994). The intent of CM is the elimination of unnecessary clinical variability, thereby lowering hospital cost and improving quality of care (Cook, 1998). Hospital costs are reduced by shortening the length of hospitalization or reducing unplanned readmissions (Renholm et al., 2002; Zander, 1988). Prolonged LOS and hospital readmission are representative of adverse outcomes that high-risk populations experience. For example, 44% of survivors of a hospitalization for congestive heart failure were readmitted to a hospital at least once within 6 months of discharge (Krumholz et al., 1997). In practice, nurse case managers work with these high-risk populations that consist of vulnerable individuals with chronic and catastrophic illness (Lamb, 1992).

Although preliminary reports evaluating the impact of CM have been positive, few research studies have focused on the effects of CM (Dechairo, 2000). A majority of the reports describing the effectiveness of CM are more descriptive and methodologically weak. The most commonly used research design is the preexperimental single group pretest, posttest design. These preexperimental designs are useful for building administrative support for CM, but they are weak designs for establishing a causal relationship between CM and patient outcomes (Lamb, 1992). Furthermore, it is unknown whether the processes related to CM have the potential to achieve desired outcomes because they have not been explored systematically (Cook, 1998). Cook investigated the effectiveness of inpatient CM on patient satisfaction, provider satisfaction, quality of care, costs, and LOS, using a meta-analysis method. However, most of the studies included in his research did not provide sufficient data and, consequently, no conclusions were made about the effect of inpatient CM on desired outcomes. Therefore, this meta-analytic study examined the effectiveness of hospital-based CM by selecting only experimental studies, which is the most powerful approach to support the conclusions. This study will provide findings regarding the effect of CM on LOS and readmission.

**Purpose, Research Question, and Variable Definitions**

The purpose of this study was to investigate the effect of hospital-based CM as compared with usual care on length of hospital stay and readmission rate by using a meta-analytic method. The research question was, “Is case management effective in reducing the hospitalization stay of inpatients and readmission rate?”

Length of hospital stay is defined as the average number of days hospitalized per patient during the individual study period. The readmission rate refers to the proportion of patients readmitted at least once within the duration of follow-up established for each study. The independent variable is any randomized implementation of CM that was performed for inpatients. Hospital-based CM is defined as a dynamic system of care involving construction of interdisciplinary protocols, continual monitoring, and facilitation of a treatment plan (Cook, 1998).

**Methods**

**Searching for Relevant Literature**

A variety of search strategies were employed to find published studies on the effectiveness of CM. The first strategy involved computerized database searches using MEDLINE (1966–2003), CINAHL (1982–2003), and HealthSTAR (1975–2003). The following key words were used: CM, care management, managed care, critical path, disease management, effect, impact, evaluation, experimental studies, randomized, and intervention. The second strategy was footnote chasing, which examines studies cited in previous reviews of CM. Third, content experts were contacted to identify any studies missed by the electronic searches.

The combination of these three strategies revealed a total of 129 articles in English-language journals. Some studies for which different outcomes were reported in more than one article were included here as a single article. The first author reviewed each of 129 abstracts.

**Selection of Studies**

The author and another reviewer independently reviewed 25 studies that reported the effect of hospital-based CM for inclusion in the meta-analysis. Studies were included if they met the following criteria: (a) sample included adults aged 18 years and more; (b) intervention was hospital-based CM for inpatients; (c) the design was randomized experimental; (d) information was provided regarding the difference in LOS or readmission rate as an outcome measure; and (e) the number of participants in the study groups was reported. Studies in which patients were mentally ill or received outpatient services were excluded. In addition, studies implementing hospital-to-community–based or community-based CM were excluded. The reviewers agreed that 12 studies met the criteria (Cline, Israelsson, Willenheimer, Broms, & Erhardt, 2000; Egan, Clavario, Burridge, Teuwen, & White, 2002; Falconer, Roth, Satin, Strasser, & Chang, 1993; Fitzgerald, Smith, Martin, Freedman, & Katz, 1994; Gagnon, Schein, McVey, & Bergman, 1999; Laramee, Levinsky, Sargent, Ross, & Callas, 2003; Naylor et al., 1999; Pugh, Havens, Xie, Robinson, & Blaha, 2001; Rich et al., 1995; Rudy et al., 1995; Schull, Tosch, & Wood, 1992; Sulch, Perez, Melbourn, & Kalra, 2000). One study employed different study designs for diagnosis-specific subgroups (Schull et al., 1992) and reviewers agreed that only subgroup data generated by a randomized experimental design would be included in analysis.
Data-Collection Methods
Each of the 12 studies was coded for author, publication year, retrieval source, country, type of CM, background of principal investigator, sample size, mean age of the study sample, diagnosis, time to outcome measure, study quality, and data to compute the effect size for LOS and number of readmissions. The first author and another coder independently abstracted data from the 12 articles to assess agreement between coders. Initial coder agreement was 95% with discrepancies between coders resolved by discussion.

The methodological and intervention quality of each study was assessed by a quality-scoring instrument consisting of two major parts. The criteria for methodological quality were adopted from a scale developed by Jadad et al. (1996), which includes a description of randomization, double blinding, and withdrawals/dropouts. The criteria for intervention quality, developed by the author, were used to assess whether the CM intervention included comprehensive assessment, education/consultation, collaboration, discharge planning, linkage with community, and monitoring/follow-up. These six components were derived from literature discussing a model of CM (Kanter, 1989; Lamb, 1992; Rothman, 1991). A score of 1 point was given for each component included in the intervention.

Using this quality rating scale, the total possible score for study quality ranged from 0 to 11. The possible range of score was then divided into three parts. Studies with scores 0–3 were considered low quality, those with 4–7 considered moderate quality, and those with 8–11 considered high quality. The authors independently rated the 12 articles for study quality, discussed those items on which they disagreed, and then reached consensus.

Statistical Analyses
Hedges’s g used for the calculation of effect size for LOS in each individual study is defined as the difference between the group means divided by the pooled standard deviation (Cooper & Hedges, 1994). A positive effect size favors the experimental group. The effect size for readmission was computed by odds ratio (OR), a relative measure of the chance of an event in the two groups (Sutton, Abrams, Jones, Sheldon, & Song, 2000). An OR less than 1.0 indicates that the CM group is less likely to have readmissions, whereas an OR greater than 1.0 means that CM group is more likely to have readmissions. A study that simply reported the results as not significant with incomplete information to compute an effect size was given the effect size of 0 as a conservative approach. When necessary, the authors contacted the corresponding author of a study to obtain the data necessary for calculation of effect size. The 95% confidence interval (CI) was determined for each study effect size and an overall mean weighted effect size was calculated weighting for study variance. The homogeneity test was performed using the Q-statistic. As the observed variance in overall mean effect size was not homogeneous, the results of the random effects model are reported here. The random effects model provides more conservative estimates of the significance of an average effect size over studies in the presence of unexplained heterogeneity (Cooper & Hedges, 1994). Sensitivity analyses were performed to examine whether the results varied by study quality, components of CM intervention, diagnoses, and country where the study was conducted.

Results
Study Selection
Of the 129 articles identified from searching strategies, 25 met initial screening criteria and were extracted for further evaluation. Of these, 13 did not meet the final inclusion criteria. One evaluated the effects of telephonic CM intervention. Five reported results in more than one publication and only one publication was selected. Two did not include the outcome of interest, three were interventions for patients under 18 years old, and two were for mentally ill patients.

Study Characteristics
In all of the studies, the control group received usual care. However, in most of the studies, what was meant by usual care was poorly defined or not defined. Eleven studies had a nurse-led CM intervention and one study had a physician-led intervention (Table 1). The studies included in this meta-analysis were of the elderly with a mean age of 72.2 years. Eight of the studies were conducted in U.S. hospital settings. Overall, the studies were of moderate to high quality, with 6 of 12 studies receiving more than 8 of 11 points. No study was assigned a low-quality rating.

The key components of the CM interventions by diagnoses are shown in Table 2. Two studies for patients with heart failure focused on intensive education programs presented by an experienced nurse (Cline et al., 1998; Rich et al., 1995). Frequently, interventions for frail elders were enforced by home visits and case manager-initiated telephone contact (Gagnon et al., 1999; Naylor et al., 1999). Four studies used their own care protocols called either a care map or a critical path (Egan et al., 2002; Falconer et al., 1993; Rudy et al., 1993; Sulch et al., 2000).

Pooled Results
Effect sizes and 95% CIs for the 12 studies are presented in Table 3. Effect sizes and 95% CI for LOS were computed using the random effects model because of heterogeneity among the studies (\( Q_T = 21.38, df = 9, p < .025 \)). The overall average weighted effect size (AWES) of LOS for 10 studies was 0.094 (\( Z = 1.46, p = .07 \)) based on \( N = 2,666 \) with a 95% CI of -0.032 to 0.220 (Figure 1). This result indicates that the CM intervention across the 10 studies was not effective in reducing hospital LOS.

The overall OR for readmission for 10 studies was 0.87 with a 95% CI of 0.69 to 1.04 (Figure 2). It can be concluded that the effect of CM on decreasing readmissions is not statistically significant at the 5% level. In terms of a Binominal Effect Size Display (Cooper & Hedges, 1994), the effect size can be interpreted as a 6% decrease in readmission for patients who received a CM intervention. No evidence of heterogeneity was found among the studies (\( Q_T = 13.24, df = 8, p > .10 \)).

Sensitivity Analyses
Length of Stay Sensitivity analyses were performed to test whether the effect size varied by quality of study, diagnoses,
and components of intervention (Figure 3). The overall AWES for five studies rated high quality was 0.119 ($Z_{H11005}/1.39, p_{H11005}/.09$) with a 95% CI of –0.049 to 0.287, while the AWES for five moderate quality studies was 0.062 ($Z_{H11005}/0.63, p_{H11005}/.27$) with a 95% CI of –0.129 to 0.253. There was homogeneity between groups ($Q_{between}/0.19, df_{H11005}/1, p_{H11022}/.10$) indicating no difference in reducing LOS between high- and moderate-quality studies.

To examine the effectiveness of interventions according to diagnoses, the studies were grouped into three major diagnoses: heart failure, stroke, and frail elder. The AWES of LOS for the heart failure group was 0.241 ($Z_{H11005}/2.059, p_{H11005}/.02$) with a 95% CI of 0.012 to 0.470. The AWES for the stroke group was –0.226 ($Z_{H11005}/–1.404, p_{H11005}/.08$) with a 95% CI of –0.542 to 0.089. The AWES for the frail elder was 0.126 ($Z_{H11005}/1.242, p_{H11005}/.11$) with a 95% CI of –0.073 to 0.324. There was no evidence of heterogeneity between groups ($Q_{between}/5.61, df_{H11005}/2, p_{H11022}/.05$). The analysis indicates that the interventions are effective in decreasing the hospitalized days of patients with heart failure whereas they are not effective for stroke patients and frail elders.

On the basis of the assumption that the implementation of CM containing all key components will result in significant differences on desired outcomes when compared with usual care, differences in effect sizes by the components of the intervention were examined. Using the intervention quality rating with a maximum six points, six studies scoring 5 or 6 were considered a high-quality intervention whereas four studies scoring fewer than 5 were considered a low-quality intervention. For the high-quality

### Table 1: Characteristics of Studies Included in Meta-Analysis

<table>
<thead>
<tr>
<th>First Author</th>
<th>Year</th>
<th>Country</th>
<th>Sample Size</th>
<th>Mean Age (years)</th>
<th>Diagnosis</th>
<th>Type of Case Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cline</td>
<td>1998</td>
<td>Sweden</td>
<td>190</td>
<td>75.6</td>
<td>Heart failure</td>
<td>Nurse-led</td>
</tr>
<tr>
<td>Egan</td>
<td>2002</td>
<td>Australia</td>
<td>66</td>
<td>67.5</td>
<td>COPD</td>
<td>Nurse-led</td>
</tr>
<tr>
<td>Falconer</td>
<td>1993</td>
<td>US</td>
<td>121</td>
<td>68.0</td>
<td>Stroke</td>
<td>Physician-led</td>
</tr>
<tr>
<td>Fitzgerald</td>
<td>1994</td>
<td>US</td>
<td>668</td>
<td>64.5</td>
<td>Frail elder</td>
<td>Nurse-led</td>
</tr>
<tr>
<td>Gagnon</td>
<td>1999</td>
<td>Canada</td>
<td>427</td>
<td>81.6</td>
<td>Frail elder</td>
<td>Nurse-led</td>
</tr>
<tr>
<td>Laramee</td>
<td>2003</td>
<td>US</td>
<td>287</td>
<td>70.7</td>
<td>Heart failure</td>
<td>Nurse-led</td>
</tr>
<tr>
<td>Naylor</td>
<td>1999</td>
<td>US</td>
<td>363</td>
<td>75.4</td>
<td>Frail elder</td>
<td>Nurse-led</td>
</tr>
<tr>
<td>Pugh</td>
<td>2001</td>
<td>US</td>
<td>58</td>
<td>74.3</td>
<td>Heart failure</td>
<td>Nurse-led</td>
</tr>
<tr>
<td>Rich</td>
<td>1995</td>
<td>US</td>
<td>282</td>
<td>79.3</td>
<td>Heart failure</td>
<td>Nurse-led</td>
</tr>
<tr>
<td>Rudy</td>
<td>1995</td>
<td>US</td>
<td>220</td>
<td>64.0</td>
<td>Critically ill</td>
<td>Nurse-led</td>
</tr>
<tr>
<td>Schull</td>
<td>1992</td>
<td>US</td>
<td>42</td>
<td>N/R</td>
<td>Epilepsy</td>
<td>Nurse-led</td>
</tr>
<tr>
<td>Sulch</td>
<td>2000</td>
<td>UK</td>
<td>152</td>
<td>74.5</td>
<td>Stroke</td>
<td>Nurse-led</td>
</tr>
</tbody>
</table>

Note: COPD = Chronic obstructive pulmonary disease; US = United States; UK = United Kingdom; N/R = not reported.

### Table 2: Key Components of Study Intervention by Diagnoses

<table>
<thead>
<tr>
<th>Components</th>
<th>Cline</th>
<th>Laramee</th>
<th>Pugh</th>
<th>Rich</th>
<th>Fitzgerald</th>
<th>Gagnon</th>
<th>Naylor</th>
<th>Stroke</th>
<th>Sulch</th>
<th>Egan</th>
<th>Rudy</th>
<th>Schull</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period of study</td>
<td>23 mo</td>
<td>N/R</td>
<td>N/R</td>
<td>4 y</td>
<td>2 y</td>
<td>3 mo</td>
<td>4 y</td>
<td>4 y</td>
<td>6 mo</td>
<td>6 wk</td>
<td>N/R</td>
<td>N/R</td>
</tr>
<tr>
<td>Assessment</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Collaboration</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Discharge plan</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Use of protocol</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Linkage</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monitoring</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Home visit</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>Phone f/u</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td>×</td>
<td></td>
</tr>
</tbody>
</table>

Note: It includes only components mentioned in the article; N/R = not reported.
The overall OR for five high-quality studies was 0.874 with 95% CI of 0.640 to 1.107, while the overall OR for five moderate-quality studies was 0.868 with 95% CI of 0.611 to 1.124 (Figure 4). The results imply no significant difference in reducing readmissions between high- and moderate-quality studies. No heterogeneity was present between two groups ($Q_{\text{between}} = 0.00$, $df = 1$, $p > .10$).

The effect sizes between groups classified by diagnosis were compared. Four studies with patients with heart failure had an OR of 0.749 with 95% CI of 0.446 to 1.052, whereas three studies with frail elders had an OR of 0.971 with 95% CI of 0.754 to 1.188. The effect of CM interventions on reducing readmission did not differ by diagnosis. There was homogeneity between groups ($Q_{\text{between}} = 1.36$, $df = 1$, $p > .05$).

The OR for six high-quality intervention studies was 0.858 (95% CI of 0.628 to 1.088), and the OR for four low-quality intervention studies was 0.887 (95% CI of 0.627 to 1.148). Neither high-quality nor low-quality intervention studies significantly reduced readmission. No evidence of heterogeneity was present between groups ($Q_{\text{between}} = 0.02$, $df = 1$, $p > .05$).

The seven studies conducted in the United States had an OR of 0.792 with a 95% CI of 0.594 to 0.989, whereas the overall OR for three studies performed outside the United States was 1.124 with a 95% CI of 0.769 to 1.478. Thus, studies conducted in the United States were effective in reducing readmission rate. No heterogeneity was found between groups ($Q_{\text{between}} = 2.57$, $df = 1$, $p > .10$).

### Publication Bias

A funnel plot is a graphical method to examine publication bias (Soeken & Sriupasanpan, 2003). The plot of variance versus effect size for LOS was shaped like a funnel (Figure 5). Although the range of variances for all but one study was narrow, this plot provides evidence that publication bias was unlikely in this meta-analysis (Cooper & Hedges, 1994).

The funnel plot for readmission was less shaped like a funnel (Figure 6). However, the fail-safe $N$s ($N_{FS}$) was 153, greater than the “reasonable guideline” value of 55.
A Meta-Analysis of the Effect of Hospital Case Management

Nursing Research  July/August 2005  Vol 54, No 4

(1984). The $N_{FS}$ is the number of unpublished studies having an average of no effect that is necessary for reducing the overall effect size from significant to non-significant (Soeken & Sripusanapan, 2003). Publication bias was unlikely to be a factor for this meta-analysis study.

Discussion

In this meta-analysis, the overall effect of hospital-based CM was not significant in reducing hospitalization days and readmission rate. As Ziguras and Stuart (2000) noted in their meta-analysis study of the effects of mental health

![FIGURE 1. Effect sizes and 95% confidence interval (CI) for length of stay.](image1)

![FIGURE 2. Odds ratio and 95% confidence interval (CI) for readmission.](image2)
CM, a possible reason for a small and nonsignificant effect may have been the differences between the CM intervention and usual care. Although the components of usual care were not clearly defined in most studies, they may have incorporated components of CM into their standard practices and procedures (Ziguras & Stuart, 2000). In addition, the effectiveness of CM may be associated with other factors, such as variation in practice or caseload among case managers, the extent of support from an institution, the organizational environment, and the patient’s relationship to the healthcare providers. With those variations, the CM model of each hospital fits the institutional environment.

According to Johnson and Schubring (1999), four changes have to occur to make a CM program successful: development of a leadership role for the case manager; development of objective-based pathways; involvement of home health in an early and sufficiently comprehensive discharge plan; and refocus of patient education process. In most of the studies included, nurses were employed as case managers to coordinate health services. Effective nurse case managers are clinical experts with management abilities and knowledge of the healthcare system (Dechairo, 2000). In addition to personal competence and knowledge, it is essential that nurse case managers have authority and responsibility to function as a central leader in facilitating and coordinating each patient’s plan of care in close collaboration with multidisciplinary team members. Nurse case managers in the studies by Naylor et al. (1999) and Rudy et al. (1995), who had a significant effect size on reducing LOS and readmission, were responsible for the process of care and accountable clinically and financially for each patient’s outcomes.

Crummer and Carter (1993) believe that the critical pathway (CP) is a hub of CM. Critical pathway involves the standardized and interdisciplinary processes that must occur for a particular type of patient to move along a continuum toward a desired outcome in a defined period of time (Renholm et al., 2002). It has been shown that the use of CP provides a reduction in the average length of hospital stay in certain diagnosis-related groups and operations (Renholm et al., 2002). Four authors of studies included in this meta-analysis indicated that they adopted CPs, which were named critical path, care pathway, intervention framework, or medical protocol. However, most of them did not describe what activities were included in CPs and how they were implemented. Johnson and Schubring (1999) developed the outcome-based pathway as a replacement for the traditional time-line type pathways. The value of outcome-based pathways is that it enables healthcare providers to focus on predefined goals and tailor their care plans to reach this predefined point.

The home care intervention has become a critical part of hospital-based CM. In many of the studies in this meta-analysis, home visits were employed by case managers. However, the frequency, intensity, and activity of home visits were not clearly stated in most studies. Home healthcare through either the case manager’s home visit or linking to the hospital’s home health agency should be specialized according to patient and caregiver needs. As implemented in Naylor’s study (1999), it may require both regularly planned and on-demand home visits to make home care successful.

One of the major components in the CM interventions was patient and caregiver education. The education
focused mainly on guidelines regarding treatment, medications, activities, and self-management associated with disease process. A good example regarding the development and implementation of education program can be found in a CM study for patients with heart failure (Cline et al., 1998). The education material was developed by a multidisciplinary team, including a physician, a nurse, and a dietitian. Next, the education program was implemented through individual visits during hospitalization by a nurse and a home visit after discharge. To reinforce the information given in the hospital, group education was held by a nurse employing oral and video presentation. The education program individualized by patient progress enables patient and family to visualize self-management of their conditions.

When sensitivity analysis was performed to examine effect sizes according to diagnosis, the CM intervention was effective for inpatients with heart failure but not for those with stroke and the frail elder. It is important to consider which patient populations are most likely to benefit from CM interventions (Lamb, 1992). The disease management programs in heart failure have been shown to be effective for cost saving and reducing the risk of hospitalization (McAlister, Lawson, Teo, & Armstrong, 2001). Cook’s (1998) systematic literature review on inpatient CM indicated that patients were more likely to have positive outcomes if their diagnoses were surgical. Outcomes such as LOS and hospital readmissions are affected by patient variables. Patient’s primary diagnosis, the number of diagnosis, and the number of surgical procedure accounted for 38% of the variance in LOS (Lave & Leinhardt, 1976). These outcomes in patients with heart failure are influenced by ischemic etiology of heart failure, left ventricular dysfunction, number of medications, and age (Babayan et al., 2003; Morrison & Beckworth, 1998). Therefore, the results here should be interpreted with caution because frail elder is not a specific diagnosis and studies targeting the frail elder may include patients with a variety of diagnoses. Some frail elders may benefit from CM and others may not.

Another methodological issue in using readmission as an outcome is the measurement of readmission. Most of the studies measured only readmissions to the study hospitals or did not clarify if readmissions to hospitals other than study hospitals were measured. Only two studies examined readmissions to hospitals other than study hospitals (Fitzgerald et al., 1994; Naylor et al., 1999).
This study has several limitations. The most obvious is the small number of studies and the inability to identify and include unpublished studies. This may result in an over- or underestimation of intervention effects. In addition, given the imprecise descriptions of the interventions and usual care, the effects of CM interventions are likely to be small.

Some may criticize the selection of LOS and readmission as outcomes. One of the optimal outcomes expected by implementation of CM is that cost saving can be achieved by decreasing the length of hospital stay. Furthermore, it may be imprecise to combine cost data derived from different calculation methods and different perspectives. Only studies using similar perspectives for cost data should be combined because each perspective has its own cost components (Centers for Disease Control and Prevention, 1992).

This meta-analysis provides information for hospital managers and researchers who are interested in hospital-based CM. Despite the many studies that reported the effectiveness of hospital-based CM, few met the inclusion criterion of a randomized trial. More randomized trials are needed to clearly define CM and usual care. Further meta-analyses are needed to test the effectiveness of CM on other outcomes.

Accepted for publication January 22, 2005.
Corresponding author: Young-Ju Kim, MSN, RN, 685 West Baltimore St, MSTF Building, Rm 5-34, Baltimore, MD 21201 (e-mail: ykim007@son.umaryland.edu).

References


---

**Erratum**

In the March/April 2005 issue of *NR*, there was an error in the article titled “A Randomized Clinical Trial of an HIV-Risk-Reduction Intervention Among Low-Income Latino Women.” In Figure 1, found on page 110, the number of participants attending only one session is 29. However, this figure (n = 29) represents the number of participants attending one or two sessions rather than just one. Therefore, this box should read as follows: “attended one or two sessions.”

This was an oversight by the author. The editor appreciates the author bringing this error to our attention.
Testing the Psychometric Properties of the Minnesota Living With Heart Failure Questionnaire

Seongkum Heo ▼ Debra K. Moser ▼ Barbara Riegel ▼ Lynne A. Hall ▼ Norma Christman

Editor’s Note
See http://www.nursing-research-editor.com for additional information provided by the authors.

► Background: Health-related quality of life (HRQOL) is an important outcome in patients with heart failure. One of the most commonly used instruments to measure HRQOL in this population is the Minnesota Living With Heart Failure Questionnaire (LHFQ). Although the psychometric properties of the LHFQ have been tested, the results do not definitively support the psychometric soundness of the instrument.

► Objective: To examine the psychometric properties of the LHFQ.

► Method: Data from 638 patients with heart failure were used to assess the reliability, homogeneity, representativeness, discriminative ability, and construct validity of the LHFQ before and after deletion of 5 items that showed lack of representativeness and contributed to inadequate factor structure.

► Results: Cronbach’s alphas for the LHFQ Total and subscales were greater than .80. Interitem correlation coefficients in 17 of the 21 items, item–total correlation coefficients in 20 items, and discriminative ability in all items were acceptable. The total and both subscales of the LHFQ differentiated New York Heart Association functional groups. The Physical subscale was moderately related to the physical measures (the Specific Activity Scale and symptom status), whereas the emotional subscale was weakly related to the measures. The results of item p level testing and factor analysis demonstrated that 7 items were consistently problematic and 5 items were recommended to be deleted. The results of the reliability, homogeneity, and construct validity after deletion of these items demonstrated that the psychometric properties of the LHFQ were improved as a result.

► Discussion: The initial results provided additional support for the reliability and substantial evidence for the validity of the LHFQ. However, the results of item and factor analyses did not fully support the psychometric soundness of several items. The psychometric properties of the LHFQ after deleting these items were improved. These results could provide researchers and clinicians a more useful measure of HRQOL.

► Key Words: health-related quality of life • heart failure • instrument

About 5 million people in the United States have heart failure, and, despite substantial advances in care and management, the number of people with heart failure is increasing each year (American Heart Association, 2003). To these patients, health-related quality of life (HRQOL) is the most important patient outcome as it reflects the impact of heart failure on their daily lives (Scott, 2000; Westlake et al., 2002). HRQOL in patients with heart failure should be accurately assessed because it is closely related to mortality and rehospitalizations (Alla et al., 2002; Konstam et al., 1996). Fundamental to accurate assessment of a patient’s HRQOL is choice of an appropriate instrument.

The Minnesota Living With Heart Failure Questionnaire (LHFQ) is the most commonly used HRQOL instrument in research on heart failure (Rector & Cohn, 1992; Rector, Kubo, & Cohn, 1987; Reddy & Dunn, 2000). A few researchers have tested comprehensively the reliability and validity of this instrument (Middel et al., 2001; Rector et al., 1987; Rector & Cohn, 1992) and demonstrated its reliability. The homogeneity of items has been tested using item–total correlations in one study, and the result was acceptable (Rector et al., 1987). However, some items were redundant in interitem correlation analysis.

Seongkum Heo, PhD, RN, is Postdoctoral Fellow; Debra K. Moser, DNSc, RN, FAAN, is Professor and Gill Chair of Nursing, College of Nursing, University of Kentucky, Lexington.
Barbara Riegel, DNSc, RN, FAAN, is Associate Professor, School of Nursing, University of Pennsylvania, Philadelphia.
Lynne A. Hall, DrPH, RN, is Assistant Dean for Research and the PhD Program and Marcia A. Dake Professor of Nursing Science; and Norma Christman, PhD, RN, FAAN, is Professor Emeritus, College of Nursing, University of Kentucky, Lexington.
(Rector et al., 1987). Several items did not constitute meaningful clusters on factor analyses (Middel et al., 2001; Rector & Cohn, 1992). Different investigators have found items to cluster in different factors (Middel et al., 2001; Rector & Cohn, 1992). Given these conflicting findings, further examination of the psychometric properties of the LHFQ is warranted.

The purpose of this study was to evaluate the following psychometric properties of the LHFQ in a large heterogeneous sample of patients with heart failure: (a) reliability (internal consistency), (b) item analyses, and (c) construct validity using factor analysis and tests of hypotheses about the relationships between the LHFQ and other measures. The following hypotheses were tested for the latter aspect of validity testing: (a) HRQOL assessed by the LHFQ will be better in patients with better functional status measured by the New York Heart Association functional classification (NYHA) and the Specific Activity Scale (SAS; Goldman, Hashimoto, Cook, & Loscalzo, 1981), (b) better HRQOL will be associated with better symptom status, and (c) better HRQOL will be associated with better health perception.

Methods

Data collected from patients with heart failure enrolled in four separate studies were used for this psychometric analysis (Moser, Macho, & Worster, 2000; Moser & Nelson, 1999; Riegel, Carlson, et al., 2002; Riegel, Carlson, Glaser, & Hoagland, 2000). The studies were three randomized, controlled trials of community-based heart failure disease management interventions and one randomized, controlled trial of a biofeedback-relaxation intervention. Participants in the four studies were similar on all baseline and clinical characteristics with the exception that patients in one site (Moser & Nelson, 1999) were younger and more were married than those in the other three sites. All data were collected using the same methods in each site. Data were collected using self-administered questionnaires and patient interviews for clinical data by trained cardiovascular nurses. The original studies were approved by the appropriate institutional review boards of the hospitals and the academic institutions. This study was reviewed and approved by the appropriate institutional review board. A total of 638 patients with heart failure were enrolled and used to examine reliability (internal consistency), item-analyses, and factor analysis. A subset of 182 patients was used for hypotheses testing to examine construct validity because this subset had data on all variables required for this testing.

Patients were recruited during a hospitalization for exacerbation of heart failure. Eligibility criteria for each of the studies were similar: (a) diagnosis of chronic heart failure associated with systolic dysfunction or preserved ejection fraction and (b) ability to read and write English or Spanish. Patients were excluded if they had unstable angina, were hospitalized for heart failure due to acute myocardial infarction, had cognitive impairments or severe psychiatric problems, or were discharged to an extended care facility.

Measurements

Health-related Quality of Life “Health-related quality of life” was defined as a patient’s subjective perception of the impact of heart failure on various aspects of his or her daily life and was measured using the Spanish and English versions of the LHFQ (Rector et al., 1987). The LHFQ was translated into Spanish by a certified company and the internal consistency was acceptable (Riegel, Carlson, Glaser, & Romero, 2003). Patients respond to the stem question, “Did your heart failure prevent you from living as you wanted during the past month by.” The LHFQ consists of 21 items that complete this stem and that are rated from 0 (no impact on HRQOL) to 5 (most negative impact on HRQOL). Higher scores indicate worse HRQOL. The total and subscale scores are obtained by adding the ratings of the relevant items. Using the LHFQ, a total score and two subscale scores (physical and emotional) are calculated. Not all items are included in the two subscale scores, but all items are included in the total scale score (Rector & Cohn, 1992). The Physical subscale consists of eight items; the Emotional subscale consists of five items. The total scale consists of all 21 items. The range of possible scores on the Physical subscale is 0 to 40, the Emotional subscale 0 to 25, and total scale 0 to 105.

Several investigators have examined the psychometric properties of the LHFQ (Bennett et al., 2002; Kubo et al., 1992; Middel et al., 2001; Ni et al., 2000; Quittan et al., 2001; Rector et al., 1987; Rector & Cohn, 1992), but only a few have comprehensively examined them (Middel et al., 2001; Quittan et al., 2001; Rector & Cohn, 1992; see http://www.nursing-research-editor.com for additional information provided by the authors). In these studies, the internal consistency of the LHFQ and homogeneity of items were acceptable (Middel et al., 2001; Quittan et al., 2001; Rector et al., 1987; Rector & Cohn, 1992). However, some items were redundant in interitem correlation analysis (Rector et al., 1987). The validity of the LHFQ has been examined using factor analysis and other types of construct validity (Middel et al., 2001; Rector & Cohn, 1992). The results of factor analysis were similar although not identical (Middel et al., 2001; Rector & Cohn, 1992). Four factors were derived, but only two factors constituted meaningful clusters. The meaningful clusters were emotional and physical factors, although the items constituting both factors were slightly different in these two studies. The items that did not cluster on these two factors were nonetheless retained because the developers believed that the items were important and had relevant effects on daily life (Rector & Cohn, 1992).

Additional construct validity has been tested using correlations between the LHFQ and other instruments or measures. The LHFQ discriminated between patients with different NYHA classes in several studies (Gorkin et al., 1993; Middel et al., 2001; Quittan et al., 2001). However, the relationships between the LHFQ and other measures including the Medical Outcomes Study Short-Form Health Survey (MOS SF-36) and MOS SF-12 were not consistent (Bennett et al., 2003; Snead, Paul, Michel, Vanbakel, & Hendrix, 2001).

Measure Used for Hypothesis Testing in Construct Validity Functional Status “Functional status” was defined as a patient’s subjective perception of his or her ability to perform...
physically demanding tasks and measured using NYHA and the SAS. Despite known difficulties with reproducibility and validity of the NYHA (Mills & Haught, 1996), it is used widely in practice and research. We included this assessment for comparison with other studies. This measure contains four classes ranging from I (no limitation of physical activity caused by cardiac disease) to IV (symptoms at rest).

The SAS is an alternative method of assessing functional status on the basis of the metabolic costs of activities (Goldman et al., 1981). The SAS contains four classes ranging from I (able to complete any activity requiring 7 metabolic equivalents) to IV (unable to complete activities requiring ≥ 2 metabolic equivalents). Activities requiring 7 metabolic equivalents include walking uphill or performing heavy outdoor work and activities requiring 2 metabolic equivalents include taking a shower or walking down 8 steps (Myers, Bader, Madhavan, & Froelicher, 2001). The reproducibility and validity of the SAS have been established by comparison with NYHA (Goldman et al., 1981).

Symptom Status “Symptom status” was defined as a patient's subjective perception of the presence and severity of abnormal physical sensations. This status was measured by assessing the presence and severity of the following symptoms: dyspnea and fatigue. Dyspnea and fatigue are the most common symptoms of heart failure (Bennett, Baker, & Huster, 1998). Patients indicated whether they had the symptom or not and the severity of the symptom. Symptom severity was rated from 1 (very mild) to 10 (worst imaginable). Scores on the symptom severity rating are summed and thus the total symptom severity score can range from 0 (no symptoms) to 20 (two symptoms as severe as imaginable). Cronbach’s alpha for this measure was .71.

Health Perception “Health perception” was defined as a patient’s assessment of his or her own overall health status. Patients’ health perception was measured using a single item, adapted from the MOS SF-36 (McHorney, Ware, & Raczek, 1993). Patients respond to the following: “Would you say your health is...?” There are five response options ranging from 1 (poor) to 5 (excellent).

Data Management and Analysis Normality, linearity, independence, and homoscedasticity assumptions were examined; there were no serious violations of the assumptions nor were there problems with multicollinearity. The internal consistency of the LHFQ was assessed using Cronbach’s alpha, and coefficients greater than .70 were considered evidence of acceptable internal consistency (Streiner & Norman, 2001).

Five different item analyses were conducted: (a) item–total correlations (Ferketich, 1991), (b) interitem correlations (Ferketich, 1991), (c) item p level (Waltz, Strickland, & Lenz, 1984), (d) discrimination index (Waltz et al., 1984), and (e) development of an item response chart (Waltz et al., 1984). Item–total and interitem correlations were conducted to examine the homogeneity of the LHFQ. An acceptable coefficient for item–total correlations is greater than .30 (meaning they contribute to the measure), and acceptable coefficients for interitem correlations are greater than .30 and less than .70 (items less than .30 = no contribution, items greater than .70 = redundant; Ferketich, 1991). The item p level test was conducted to examine the representativeness of each item. Acceptable p levels are greater than .30 and less than .70 (items less than .30 or greater than .70 are considered either poorly representative because either very few patients have that concern or most patients have that concern at the same level and thus there is no variability; Waltz et al., 1984). The discrimination index was calculated and item response chart were developed to examine the discriminative ability of each item of the LHFQ. A value of greater than .20 on the discrimination index test and a significant p value for the chi-square test for each item in the item response chart test indicate that the item has the ability to discriminate those with poor HRQOL from those with good HRQOL (Waltz et al., 1984).

Factor analysis was done using common factor analysis (principal axis factoring) with direct oblimin rotation because some correlations between factors were assumed (Pett, Lackey, & Sullivan, 2003). In the current study, the correlation coefficients between factors I and II, I and III, and II and III before the deletion of five items were .64, .22, and .24, respectively. The correlation coefficient between factors I and II after the deletion was .65. Factors were extracted on the basis of the results of a scree plot, the eigenvalues, total variance, and conceptual considerations (Pett et al., 2003; Polit & Hungler, 1999). We used a loading score of greater than .40 as a cutting point (Pett et al., 2003; Polit & Hungler, 1999). The pattern matrix was presented because this showed more reasonable clusters than the structure matrix. Additional construct validity was examined using mean differences between groups and correlation analyses. The LHFQ scores in NYHA II versus NYHA III/IV were compared using t test. The relationships between the LHFQ scores and the following variables were examined using Spearman rank correlation (p) to assess construct validity: specific activity, symptom status, and health perception. On the basis of the initial test results, five items were deleted, and all tests except item p level were repeated on the scale with these five items deleted.

Results

Initial Psychometric Properties of the LHFQ

Sample characteristics are described in Table 1.

Reliability Descriptive statistics and Cronbach’s alphas for the LHFQ are presented in Table 2. Cronbach’s alphas for the LHFQ ranged from .85 to .91, indicating adequate internal consistency.

Item Analyses Interitem correlations and item–total correlations. In the interitem correlations, the correlation coefficient between away from home difficult and relating to or doing things with friends or family difficult was .70, indicating redundancy. The coefficient between fatigue and shortness of breath was .69. The coefficients between medical costs and all other items, and between side effects from medications and all other items, were less than .30, indicating no contribution to this measure. The corrected
item–total correlation coefficients of these two items were near .30 (.3017 and .3262, respectively). Corrected item–total correlation coefficients for 20 of the 21 items were greater than .30, indicating adequate homogeneity of the items. The corrected correlation coefficient of one item, making you stay in a hospital, was .27, indicating it did not contribute this measure.

**Item p Level** In the item p level test, seven items demonstrated a value less than .30, indicating that the majority of patients did not perceive that these factors seriously affected their daily lives. These items were the following: swelling in ankles, legs, working to earn a living difficult; sexual activities difficult; medical costs; side effects from medications; feeling burden to family or friends; and difficulty concentrating or remembering.

**Discrimination Index and Item Response Chart** The coefficients of the discrimination index for all items of the LHFQ were greater than .20. The results of the chi-square tests using the item response chart were significant for all the items. These findings indicate that all items demonstrated acceptable ability to discriminate patients with extremes of HRQOL.

**Construct Validity: Factor Analysis** Three factors were extracted (Table 3). This three-factor structure explained 45% of the total item variance. Seventeen of the 21 items demonstrated moderate or strong loadings (> .40). Four items, eating less foods I like, hospitalization, medical costs, and side effects from medications, failed to load on any factor. The items that loaded on factor 1 were very similar to the items defined as those constituting the Physical subscale of the LHFQ by Rector and Cohn (1992). One item, swelling in ankles, legs, additionally loaded on this factor. The items that loaded on factor 2 were the same as those identified as constituting the Emotional subscale by the developers (1992). The items that loaded on factor 3 were those related to sexual activities difficult and working to earn a living difficult. The responses to the two items were not distributed normally. Sixty percent of the patients reported that they did not experience working to earn a living difficult and sexual activities difficult. The item, recreational activities difficult, double loaded on factors 1 and 3.

**Other Construct Validity: Correlations Between LHFQ Total and Subscales Scores and NYHA, SAS, Symptom Status, and Health Perception** To examine construct validity, several hypotheses were tested, and all hypotheses were supported (Tables 4 and 5). The mean scores of the total and subscales of the LHFQ in the group with NYHA II were lower than those of the other groups, indicating that the LHFQ is sensitive to differences in HRQOL among patients with NYHA II.
(indicating better HRQOL) than those in the group with NYHA III–IV. The SAS level was positively moderately related to the total and Physical subscales of the LHFQ and was positively weakly related to the Emotional subscale. Patients with better functional status had better HRQOL. Symptoms (dyspnea and fatigue) were more strongly related to the Physical subscale than the Emotional subscale. Patients with more severe symptoms had worse HRQOL. The total and subscales of the LHFQ were negatively related to health perception. Patients with more negative perceptions of their own health had poorer HRQOL.

A number of demographic and clinical characteristics are known to be associated with HRQOL (Grady et al., 1995) and thus could confound the associations demonstrated in the construct validity tests. To control for this potential confounding influence, the hypotheses above were tested again, controlling for age, gender, heart failure etiology, and employment status. The findings were unchanged by the addition of these covariates.
Psychometric Properties After Deleting Five Items
The psychometric properties of the LHFQ after deleting the following five items were examined because these items showed lack of representativeness and factor structure. The items were working to earn a living difficult, sexual activities difficult, hospitalization, medical costs, and side effects from medications.

Reliability and Item Analyses Internal consistency as demonstrated by Cronbach’s alpha for the total scale was .92. The item–total correlation coefficients of the 16 items were greater than .30, and interitem correlation coefficients of all items were greater than .30 and less than .70, indicating acceptable homogeneity of each item. The coefficients of the discrimination index were greater than .20, and p values of the item response chart were significant in all items, indicating acceptable ability to discriminate patients with extremes of HRQOL by these two criteria.

Construct Validity: Factor Analysis Two factors were extracted on the basis of a scree plot, eigenvalue result, and theoretical considerations. The two factors explained 50% of the total variance. Fifteen of the 16 items demonstrated moderate or strong loadings (> .40), and only one item, eating less foods I like, did not load on any factor (Table 3).

Other Construct Validity: Correlations Between the LHFQ Total Scale and NYHA, SAS, Symptom Status, and Health Perception The new LHFQ Total score discriminated NYHA groups (NYHA II vs. III/IV). The total score was moderately correlated with the SAS, symptom status, and health perception (Table 5).

Discussion
Although the LHFQ is the most commonly used HRQOL instrument in research on heart failure, prior studies in which the psychometric properties of the LHFQ were tested indicate that the instrument may lack sensitivity (Riegel, Moser, et al., 2002) and has some problems with its factor structure and the homogeneity of items (Middel et al., 2001; Rector et al., 1987; Rector & Cohn, 1992). The current study supports the internal consistency of the LHFQ and provides some additional support for the construct validity of the LHFQ. However, the results of our item analyses and factor analysis suggest that at least two items need to be reworded and five items deleted to improve the psychometric properties of the LHFQ. We again tested the psychometric properties of the instrument with the five items deleted and demonstrated improvement with the resulting 16-item instrument.

Initial Psychometric Properties of the LHFQ
The results of the reliability tests of each of the LHFQ Total and subscales were similar to those reported in other studies (Middel et al., 2001; Quittan et al., 2001; Rector & Cohn, 1992). In a number of studies, including the current study, Cronbach’s alphas for the total and subscales of the LHFQ were greater than .80, demonstrating consistently acceptable reliability.

The interitem and corrected item–total correlations that test the homogeneity of the items of the LHFQ within a construct have rarely been discussed in the literature. Rector et al. (1987) examined these properties and seven items were considered redundant. In the current study, only two items were considered redundant. The current study had a large heterogeneous sample, so the results may be more reliable. Thus, data from our study indicate that there is no serious redundancy among the items.

The results of corrected item–total correlations in the current study supported the prior study (Rector et al., 1987). In the current study, one item did not contribute substantially to this measure. The coefficients of interitem correlations for items medical costs and side effects from medications were low in the current study, and the coefficient of corrected item–total correlations for item side effects from medications was relatively low in the Rector et al. study (1987). This means that the latter item did not contribute importantly to this measure. Thus, considerations of other psychometric properties of these two items are needed.

The representativeness and discriminative ability of individual items of the LHFQ have not been examined previously. The results of the item p level test showed that seven items did not contribute to the LHFQ scores. Among these seven items, the following five items did not consistently belong to any meaningful factor in the factor analysis: swelling in ankles, legs; working to earn a living difficult; sexual activities difficult; medical costs; and side effects...
The results of the factor analysis in the current study provided support for the two factors proposed by Rector and Cohn (1992). The items that loaded on both factors in the current study were nearly identical to those identified by Rector and Cohn (1992). Two items loading on factor 3 were exactly the same as the items loaded on factor 3 in Rector and Cohn (1992). In Rector and Cohn (1992), factor 3 did not constitute a separate factor due to abnormal distribution of the items belonging to the factor. The results of the current study supported this conclusion.

Factor analysis and item analyses in this study and other studies revealed that there are some problematic items on the LHFQ. The item "swelling in ankles, legs" loaded on different factors in different studies. In this study, the item loaded on factor 1 defined as the Physical subscale by Rector and Cohn (1992). However, this item double loaded on factors 1 and 2 in Middel et al. (2001) and it loaded on factor 4 (no meaningful cluster) in Rector and Cohn (1992). In a qualitative study (Hak, Willems, van der Wal, & Visser, 2004), respondents were confused by this item because some had swelling in legs but not in their ankles. The item "recreational activities difficult" double loaded in the current study and the prior studies (Middel et al., 2001; Rector & Cohn, 1992). Of the two factors, one was the factor identified as the Physical subscale by Rector and Cohn (1992). The Physical subscale of the LHFQ consists of items related to physical activities and symptoms, and these two items are related to physical activity or symptom. Thus, refinements of these items and further factor analysis may be needed to support construct validity and justify the subscales of this instrument.

The items "working to earn a living difficult", "sexual activities difficult", and "side effects from medications" loaded on factor 3 or 4 or did not load on any factor in this study nor in prior studies (Middel et al., 2001; Rector & Cohn, 1992). None of these three items constituted meaningful factors, and all of these items had low p levels in the item p level test, indicating most patients did not perceive that these conditions seriously affected their daily lives. The items "hospitalization" and "medical costs" did not load on any factor or loaded on different factors in different studies including the current study (Middel et al., 2001; Rector & Cohn, 1992). In the current study, the item-total correlation coefficient or item p level of these two items was low. Although it seems reasonable that problems with these items could affect HRQOL, few researchers have examined these items in relation to HRQOL. The results of the current study and prior study (Bennett et al., 1998) indicate that these problems were not major concerns for the majority of patients. Thus, their inclusion in the LHFQ may decrease the measure’s validity.

In this study, construct validity was additionally assessed through a series of hypotheses and the results provided support for the construct validity of the LHFQ. Similar to other studies, LHFQ scores discriminated between patients in two NYHA classes (Gorkin et al., 1993; Middel et al., 2001; Quittan et al., 2001). These findings provide support for construct validity of the LHFQ. In this study, health perception was relatively strongly related to the LHFQ, consistent with the few investigators who have examined this relationship (Grady et al., 1995), and providing further support for construct validity.

Psychometric Properties of the LHFQ After Deleting Five Items

The Cronbach’s alpha coefficient for the total scale of the LHFQ increased by .01 after deleting five items. Cronbach’s alpha is affected by not only the magnitude of correlations of the items of a measure, but also the number of items (Streiner & Norman, 2001). That is, if items are deleted from a reliable measure, Cronbach’s alpha may decrease. However, the reliability of the LHFQ remained acceptable even after deletion of five items. The results of item-analyses indicated that homogeneity and discrimination ability of all items were acceptable even after deleting the items.

In factor analysis, all items except the item "eating less foods I like" loaded on two factors, and the factors constituted meaningful clusters: physical and emotional. These two factors explained more variance than the three factors in the initial analysis. The correlation coefficients between the LHFQ new total scale and the SAS and symptom status were slightly stronger than those before deleting the items. The correlation coefficient between the LHFQ new total scale and health perception was similar to those before deleting the items. The results of these tests demonstrated that the construct validity of the LHFQ was improved after deleting the items.

The results of this study provide support for the reliability of the LHFQ and suggest changes that could improve the psychometric properties of this measure. The internal consistency estimates for all total and subscales were acceptable. The results of construct validity tests provide substantial evidence for the construct validity of the LHFQ. However, the results of item analyses and factor analysis demonstrated problems in specific areas that should be addressed to improve the psychometric properties of the LHFQ. On the basis of these results, the psychometric tests were conducted again after deleting the following five items: "working to earn a living difficult", "sexual activities difficult", "hospitalization", "medical costs", and "side effects from medications". The results of the retests demonstrated that the psychometric properties of the LHFQ were better than before deleting the items. On the basis of the results of item analyses and factor analysis, we recommend that the following items be reworded to improve the psychometric properties of the LHFQ: "swelling in ankles, legs" and "recreational activities difficult". Such changes could improve the reliability and validity of the LHFQ and provide researchers and clinicians with an even more useful measure of HRQOL.

Accepted for publication February 28, 2005.

Corresponding author: Debra K. Moser, DNSc, RN, FAAN, College of Nursing, University of Kentucky, 527 CON Bldg, 760 Rose St, Lexington, KY 40536 (e-mail: dmoser@uky.edu).
References


Predicting Nosocomial Bloodstream Infections Using Surrogate Markers of Injury Severity
Clinical and Methodological Perspectives

Maher M. El-Masri  ▼  Tarek A. Hammad  ▼  Susan M. Fox-Wasylyshyn

Background: Injury severity indices are numerical scores that are utilized to predict nosocomial bloodstream infections (BSI) in critically ill patients. However, surrogate markers of injury severity (SMIS) may be more clinically meaningful than these commonly used numerical injury severity indices with respect to the control and prevention of nosocomial BSI.

Objective: The purpose of this study was to demonstrate the clinical and research implications of using the SMIS in predicting nosocomial BSI.

Method: A prospective nonexperimental cohort study was conducted on 361 critically ill trauma patients. Three logistic regression models were examined for their clinical relevance and statistical parsimony. The first model included the Injury Severity Score (ISS) and 5 other independent predictors, and excluded the SMIS. The second model included all study variables. The third model excluded the ISS.

Results: The analysis suggested that number of blood units transfused, number of central venous catheters inserted, and use of chest tube(s) were the SMIS. The ISS was found to be an independent predictor of nosocomial BSI only when the SMIS were not included in the model. The model that included the SMIS and excluded the ISS explained the highest variance in nosocomial BSI and had the best negative predictive value (93%).

Discussion: Clinicians can use knowledge of SMIS to develop interventions that minimize the risk of nosocomial BSI. Hence, the SMIS can serve not only as a prediction tool but also as a way to enhance control and prevention strategies for BSI.

Key Words: injury severity · nosocomial bloodstream infection · surrogate markers · trauma patients

The Injury Severity Score (ISS) is a numerical index that reflects the degree of anatomic injury in trauma patients and has been utilized to predict nosocomial bloodstream infections (BSI) among the critically ill (El-Masri, Joshi, Hebden, & Korniewicz, 2002; Hurr, Hawley, Czachor, Markert, & McCarthy, 1999). Although the ISS provides a tool for the prediction of nosocomial BSI, it has little clinical value in terms of prevention and management of nosocomial BSI. Therefore, replacing numerical injury severity indices with more clinically relevant surrogate markers of injury severity (SMIS) can provide clinicians with specific targets on which they can intervene to minimize the risk of nosocomial BSI. To date, however, no studies have examined the research and clinical significance of the SMIS in the prediction of nosocomial BSI. Hence, this study examines the methodological and clinical usefulness of the SMIS as an alternative that extends beyond the mere predictive value of the ISS.

Background
Trauma is a leading cause of death in the United States (American College of Surgeons Committee on Trauma, 1999; Centers for Disease Control and Prevention [CDC], 2001; Mokdad, Marks, Stroup, & Gerberding, 2004) and Canada (Sampalis, 2002). The impact of physical trauma on the American healthcare system exceeds $83 billion per year (CDC, 1995). In fact, costs of workplace-related injuries to industry and society at large are estimated to exceed $121 billion per year (CDC, 1999). In Canada, trauma accounts for 13,000 deaths and 2 million hospital days annually. The total direct and indirect costs of trauma care to the Canadian healthcare system exceeds $14 billion per year, representing 11% of total healthcare expenditures (Sampalis, 2002).

Despite advancements in trauma care and resuscitative management that have increased the immediate survival of critically ill trauma patients, they have a particularly high risk for developing nosocomial infections (Erbay et al., 2003; Wallace, Cinat, Gornick, Lekawa, & Wilson, 1999). This risk might be attributed to the aggressive management strategies that are frequently used with these patients and to their unique case presentation. Trauma patients present with unclean severe multiple injuries, undergo multiple aggressive therapeutic measures that include a wide range of invasive procedures, and have multiple portals for bacteria. In addition,
critically ill trauma patients are often hospitalized for prolonged periods of time and sustain inadequate nutritional support (Bochicchio, Joshi, Knorr, & Scalea, 2001; Papia et al., 1999; Pittet, Tarara, & Wenzel, 1994).

Although all nosocomial infections pose a threat to trauma patients, nosocomial BSI are especially detrimental. The incidence of nosocomial BSI was reported to be second only to pneumonia among critically ill adult trauma patients (Hurr et al., 1999). Consequences of nosocomial infections in general and nosocomial BSI in particular include health-related complications, increased mortality, and increased hospital length of stay (Digiavine, Chenoweth, Watts, & Higgins, 1999; Karchmer, 2000; Pittet et al., 1994; Saint, Veenstra, & Lipsky, 2000; Wenzel, 1995). In fact, nosocomial BSI are responsible for 3.5 million additional hospital days per year in the United States and $3.5 billion in related costs (Pittet, 1997). The annual nationwide expenditure for the actual treatment of nosocomial BSI has been reported to be well over $8.5 million (Keita-Perse & Gaynes, 1996; Wenzel, 1997).

Although the predictors of nosocomial BSI have been studied extensively in nontrauma populations (Harbarth, Samore, Lichtenberg, & Carmeli, 2000; Pittet, Davis, Li, & Wenzel, 1997; Rojo, Pinedo, Clavijo, Garcia-Rodriguez, & Garcia, 1999; Ryan et al., 1997), few studies have examined predictors of these infections among critically ill trauma patients (Antonelli et al., 1996; El-Masri et al., 2002; El-Masri, Hammad, McLeskey, Joshi, & Korniewicz, 2004). El-Masri et al. (2002) conducted a case-control retrospective study with a group of critically ill trauma patients and found that the ISS was a significant predictor of nosocomial BSI. Antonelli et al. (1996) reported that the risk of late-onset BSI among trauma patients was significantly higher among patients with a high abbreviated injury score (the main component of the ISS).

Contrary to the aforementioned studies, in a later study, El-Masri et al. (2004) found that the ISS was not an independent predictor of nosocomial BSI. Interestingly, El-Masri’s study examined the potential confounding influence of many more variables than was examined in the previous studies. It was therefore hypothesized that the failure to find an association between the ISS and nosocomial BSI in El-Masri’s study was attributed to the inclusion of other variables that served as the SMIS.

**Purpose**

The purpose of this study was to examine the impact of the SMIS on the relationship between the ISS and nosocomial BSI. Findings of the study are expected to demonstrate the clinical and research implications of using the SMIS in predicting nosocomial BSI. Replacing numerical injury severity indices with more clinically relevant surrogate markers as predictors of nosocomial BSI can provide clinicians with more useful information with which to intervene in an effort to minimize the risk of these infections. In other words, unlike injury severity indices, surrogate marker variables can serve not only as prediction tools but can also inform intervention strategies for prevention and control of nosocomial BSI.

**Methods**

**Design**

A prospective nonexperimental cohort study was conducted at a mid-Atlantic coast shock trauma center. Four hundred twenty-seven patients admitted to two critical care trauma units were screened for eligibility over an 8-month period. Sixty-six patients did not meet the inclusion criteria and were excluded from the study, producing a sample of 361 subjects. Patients who qualified for enrollment in the study were aged 18 years and above, had sustained a critical traumatic injury to at least one body system, and were admitted to a critical care trauma unit (intensive care unit [ICU]). Patients experiencing a second or subsequent episode of nosocomial BSI during their current hospitalization were excluded. Patients were excluded also if they were transferred from the ICU or died within 48 hr of admission to ICU, or if they were direct admissions to a step-down ICU. The study population was managed according to the standards of care implemented by the institution. According to these standards, decisions regarding procedures such as changing of central venous catheters and treatment decisions such as administration of prophylactic antibiotics were made on a case-by-case basis.

**Data Collection**

Patients’ medical records were reviewed every 48 to 72 hr. Data collection was deemed to be complete upon positive diagnosis of nosocomial BSI or discharge from the ICU without nosocomial BSI. When patients were discharged with blood or catheter-tip culture results pending, the results were followed and included in the analysis.

**Variable Definitions**

An extensive review of the literature pertaining to variables associated with the development of nosocomial BSI resulted in the identification of 43 variables listed in Table 1. Only nine variables found to be independent predictors of nosocomial BSI were used in the model-building procedures described in this study.

According to the CDC criteria for the definition of nosocomial infections, a *nosocomial BSI* was defined as a laboratory-confirmed positive blood culture for bacteria, fungus, or both that was collected more than 48 hr after admission (Garner, Jarvis, Emori, Horan, & Hughes, 1988). The study end-point was either diagnosis with nosocomial BSI or discharge from the ICU without nosocomial BSI. Total number of blood units transfused and central venous catheters inserted were measured as the total number from admission until the study end-point. Percent change in serum albumin was calculated as the percent of the difference between the admission albumin value and the final value at the study end-point divided by admission value, and was coded to reflect the percent change per 10% increment.
Outcome length of stay was defined as the length of stay before the study end-point. The variable patient disposition reflected whether the patient was admitted directly from the scene of trauma or was transferred from another healthcare institution. Microbial resistance was defined as the presence of methicillin-resistant Staphylococcus aureus or vancomycin-resistant Enterococci as documented by an infectious diseases specialist and culture. A preexisting infection was defined as any confirmed infection that occurred between admission and the study end-point.

Severity of injury was measured in terms of the patient’s ISS (Baker & O’Neill, 1976). The ISS is an anatomic injury severity score and is calculated by summing the squares of the three highest Abbreviated Injury Severity (AIS) scores obtained over six body regions (head and neck, face, thorax, abdomen, extremities, and external). The score for any of the six AIS body regions ranges from 1 (minor injury) to 5 (fatal injury). The ISS therefore ranges from 0 to 75, with a score of 0 indicating no injury and a score of 75 indicating a nonsurvivable fatal injury. The ISS takes approximately 30 min to complete and is generally calculated retrospectively by trained personnel after all injuries have been diagnosed (Civil & Schwab, 1988). Like all injury severity indices, the psychometric properties of the ISS have been tested only in terms of its predictive validity with respect to outcomes such as mortality and hospital length of stay (Linn, 1995).

Data Analysis
Data were screened and appropriately treated for missing data, outliers, multicollinearity, singularity, and deviation from normality. Univariate data analyses that included general frequencies, means and standard errors, χ², and student’s t test were performed to examine the crude associations between the outcome and each of the study variables. Only variables that had significant associations with the outcome at a liberal α ≤ .25 (Table 1) were retained for use in the regression analysis (Hosmer & Lemeshow, 2001). Stepwise regression analysis identified the nine independent predictors of nosocomial BSI used in this report (Table 3). Additional details on the univariate and regression analysis procedures that led to the identification of these independent predictors are available (El-Masri et al., 2004).

To identify the SMIS, a linear regression model was built in which the ISS was treated as a dependent variable to examine its predictor variables (Table 2). Then, three logistic regression models (Table 3, Models 1–3) were built to assess the impact of the identified SMIS on the relationship between the ISS and nosocomial BSI and to examine which model provides the most clinical relevance and statistical parsimony. The first model included the ISS and five other independent predictors, and excluded the SMIS. The second model included all study variables. The third model excluded the ISS only.

A 95% confidence interval that did not include one was set as the criterion to establish significance for the individual variables. The predictive accuracy as reflected by various classification indices (specificity, sensitivity, positive predictive value, and negative predictive value) was measured for each regression model. The proportion of variance explained by the covariates (R²) was examined also.

Results
Table 2 shows the standardized regression coefficients of the linear regression model that was built to examine which of the nine study variables predict the ISS. The data demonstrate that predictors of the ISS were number of central venous catheters, number of blood units transfused, and presence of a chest tube(s).

Table 3 presents the crude and adjusted odds ratios (ORs) of the association between the study variables and nosocomial BSI. The adjusted ORs in Table 3 are presented in three logistic regression models (Models 1–3) that demonstrate the impact of the presence of chest tube(s), the number of blood units transfused, and the number of central venous catheters inserted on the relationship between the ISS and nosocomial BSI. All three logistic regression models included in the analysis had appropriate data fit, as indicated by the nonsignificant Hosmer-Lemeshow statistics for model goodness of fit (Table 3). The crude analysis shows that the ISS has a significant association with nosocomial BSI. The OR of 1.49 for the ISS indicates that for each 10-unit increase in the ISS, there is a 49% increase in the odds of developing nosocomial BSI. Model 1 demonstrates that the ISS was an independent predictor of nosocomial BSI (OR = 1.68; 95% CI, 1.75–2.21) when its three presumed surrogate markers were not included in the analysis. Adjusting for patient disposition, change in serum albumin level, outcome length of stay, use of immunosuppressive drugs, microbial resistance, and presence of other preexisting infections increases the odds of nosocomial BSI to 68% as compared to 49% in the crude analysis. However, Model 2 demonstrates that when the number of blood units transfused, the number of central venous catheters inserted, and the presence of a chest tube(s) are included in the analysis, the ISS is no longer an independent predictor of nosocomial BSI (OR = 1.22; 95% CI, 0.90–1.66).

The results in Table 3 suggest that the exclusion of ISS, while accounting for its surrogate markers, yields a regression model that explains a comparable percentage of variance (Cox & Snell R² = .313, Nagelkerke R² = .544) to that explained in Model 2 (Cox & Snell R² = .315, Nagelkerke R² = .549), but superior to that explained in Model 1 (Cox & Snell R² = .142, Nagelkerke R² = .247).

Table 4 presents the classification indices (specificity, sensitivity, positive predictive value, and negative predictive value) for each of the three regression models. The data suggest that Model 3 has specificity comparable to the other two models. However, it has the best sensitivity (60%), positive predictive value (76.7%), and negative predictive value (93%) of the three models.
TABLE I. Variables Considered in the Univariate Analysis

<table>
<thead>
<tr>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean arterial BP</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Received antibiotics</td>
</tr>
<tr>
<td>Glasgow coma scale</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Received prophylactic antibiotics</td>
</tr>
<tr>
<td>Injury severity score</td>
</tr>
<tr>
<td>Race</td>
</tr>
<tr>
<td>Number of antibiotics</td>
</tr>
<tr>
<td>Initial serum albumin level</td>
</tr>
<tr>
<td>IV drug abuse</td>
</tr>
<tr>
<td>Initial WBC count</td>
</tr>
<tr>
<td>Percent change of serum albumin</td>
</tr>
<tr>
<td>Comorbidity</td>
</tr>
<tr>
<td>Final WBC count</td>
</tr>
<tr>
<td>Outcome ICU length of stay</td>
</tr>
<tr>
<td>Number of lumens in jugular line</td>
</tr>
<tr>
<td>Percent change in WBC count</td>
</tr>
<tr>
<td>Total ICU length of stay</td>
</tr>
<tr>
<td>Presence of microbial resistance</td>
</tr>
<tr>
<td>Number of lumens in femoral line</td>
</tr>
<tr>
<td>Presence of surgical drainage</td>
</tr>
<tr>
<td>Presence of preexisting infection</td>
</tr>
<tr>
<td>Number of intravascular catheters</td>
</tr>
<tr>
<td>Number of lumens in subclavian line</td>
</tr>
<tr>
<td>Type of trauma</td>
</tr>
<tr>
<td>Central venous catheter days</td>
</tr>
<tr>
<td>Use of peripheral IV</td>
</tr>
<tr>
<td>Nurse-patient ratio</td>
</tr>
<tr>
<td>Number of blood transfusion</td>
</tr>
<tr>
<td>Use of peripheral arterial line</td>
</tr>
<tr>
<td>Type of injury</td>
</tr>
<tr>
<td>Colloid administration</td>
</tr>
<tr>
<td>Use of central arterial line</td>
</tr>
<tr>
<td>Mechanism of injury</td>
</tr>
<tr>
<td>Had surgery during ICU stay</td>
</tr>
<tr>
<td>Presence of femoral line</td>
</tr>
<tr>
<td>Patient disposition</td>
</tr>
<tr>
<td>Type of surgery</td>
</tr>
<tr>
<td>Presence of subclavian line</td>
</tr>
<tr>
<td>Received CVVH</td>
</tr>
<tr>
<td>Received immunosuppressive drugs</td>
</tr>
<tr>
<td>Presence of internal jugular line</td>
</tr>
<tr>
<td>Received TPN</td>
</tr>
<tr>
<td>Presence of chest tube</td>
</tr>
</tbody>
</table>

Note: BP = blood pressure; IV = intravenous; WBC = white blood cells; ICU = intensive care unit; CVVH = continuous venous to venous hemodialysis; and TPN = total parenteral nutrition.

*p > .25; variable therefore was not considered in the regression analysis.

Discussion

The ISS is widely used in trauma research to assess the extent of anatomic trauma injuries. It also has been used to predict the risk of developing nosocomial infections among different patient populations (Baker, Meredith, & Haponik, 1996; El-Masri et al., 2002; Pories et al., 1991). In this study, the clinical relevance of ISS in predicting nosocomial BSI among critically ill trauma patients was examined when the SMIS was taken into account. Our data suggested that number of blood units transfused, number of central venous catheters inserted, and presence of chest tube(s) were all independent predictors of the ISS, and could therefore be considered to be the SMIS. The data suggested that when the SMIS replace the ISS in a nosocomial BSI predictive model, the predictive accuracy of the model is enhanced. This finding carries important research and clinical implications that should be carefully considered in predictive modeling studies.

The findings demonstrate that when building predictive models, it is important to consider the potential impact of all variables that may influence the relationship between the predictors of interest and the outcome variable. A predictive model that best represents an outcome of interest is driven by clinical utility, and is one in which investigators use their clinical and theoretical knowledge to identify the best combination of variables to be investigated. Failure to do so may obscure clinically relevant information. In fact, this study accounted for variables such as presence of a chest tube(s) and number of blood units transfused as potential independent predictors of nosocomial BSI and for as many as 45 variables that were identified in the literature. It is, therefore, not surprising that our findings were incongruent with those of previous studies (Agarwal, Murphy, Cayten, & Stahl, 1993; El-Masri et al., 2002; Pories et al., 1991) that suggested an independent association between the ISS and nosocomial BSI because these studies were often limited by the number and type of variables analyzed.

When an investigator builds a predictive model that includes two or more variables having a collinear relationship, the fit of the model might be compromised due to the redundancy of data, which leads to inflation of error terms (Nunnally & Bernstein, 1994; Tabachnick & Fidell, 2001). Therefore, when collinearity is an issue among a set of variables, it is important that the clinical relevance of these variables be considered when deciding which variables to include. The decision should not be an arbitrary one, but should be based on theoretical knowledge and the clinical implications of each variable as it relates to the outcome variable. In our study, Model 2 is largely comparable to Model 3 in terms of the variance explained (Cox & Snell $R^2 = .315$ and .313, respectively) and fit of the model. However, Model 3 has more clinically relevant variables than Model 2.

A clinically relevant predictive model should provide clinicians with efficient classification indices (i.e., specificity, sensitivity, positive predictive value, and negative predictive values). These indices are tools that allow clinicians to determine the clinical value of a predictive model in terms of its ability to accurately classify patients with respect to an outcome variable. In our study, model
### TABLE 2. Standardized Regression Coefficients of the Injury Severity Score (ISS) Linear Regression Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>β</th>
<th>SE</th>
<th>T</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of central venous catheters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4–5 lines</td>
<td>.180</td>
<td>1.92</td>
<td>2.389</td>
<td>.017*</td>
</tr>
<tr>
<td>6 or more lines</td>
<td>.220</td>
<td>2.295</td>
<td>2.739</td>
<td>.007*</td>
</tr>
<tr>
<td>Number of blood transfusion units</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4–9</td>
<td>.034</td>
<td>1.513</td>
<td>0.502</td>
<td>.616</td>
</tr>
<tr>
<td>10 or more</td>
<td>.187</td>
<td>1.678</td>
<td>2.47</td>
<td>.011*</td>
</tr>
<tr>
<td>Presence of chest tube(s)</td>
<td>.175</td>
<td>1.428</td>
<td>3.289</td>
<td>.001*</td>
</tr>
<tr>
<td>Serum albumin percent change</td>
<td>-.015</td>
<td>0.028</td>
<td>−.283</td>
<td>.777</td>
</tr>
<tr>
<td>Outcome length of stay</td>
<td>-.104</td>
<td>0.107</td>
<td>−1.560</td>
<td>.120</td>
</tr>
<tr>
<td>Used immunosuppressive</td>
<td>.003</td>
<td>2.576</td>
<td>0.063</td>
<td>.950</td>
</tr>
<tr>
<td>Had microbial resistance</td>
<td>.030</td>
<td>3.153</td>
<td>0.557</td>
<td>.578</td>
</tr>
<tr>
<td>Had other pre-existing infection(s)</td>
<td>.088</td>
<td>1.476</td>
<td>1.471</td>
<td>.142</td>
</tr>
<tr>
<td>Disposition</td>
<td>-.004</td>
<td>1.805</td>
<td>−0.081</td>
<td>.935</td>
</tr>
</tbody>
</table>

Note. ISS = dependent variable; β = standardized regression coefficient.  
*Significant regression coefficient at an α of .05.

### TABLE 3. Crude and Adjusted Odds Ratio of Nosocomial Bloodstream Infections Across Different Regression Models

<table>
<thead>
<tr>
<th>Variable</th>
<th>Crude OR (95% CI)</th>
<th>Adjusted OR (95% CI)</th>
<th>Model 1*</th>
<th>Model 2b</th>
<th>Model 3c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Had a chest tube</td>
<td>4.0 (2.20–7.27)*</td>
<td>3.86 (1.65–9.05)*</td>
<td>4.32 (1.80–9.95)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of blood units transfused</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 or more</td>
<td>7.73 (4.16–14.36)*</td>
<td>9.93 (1.79–55.03)*</td>
<td>5.16 (2.26–11.75)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–9</td>
<td>1.0 (reference)</td>
<td>1.0 (reference)</td>
<td>1.0 (reference)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of CVCs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 or more</td>
<td>9.90 (5.26–18.63)*</td>
<td>38.29 (3.54–413.72)*</td>
<td>22.11 (6.56–74.47)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4–5</td>
<td>0.83 (0.44–1.58)*</td>
<td>6.04 (0.61–60.33)</td>
<td>3.25 (1.07–9.87)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0–3</td>
<td>1.0 (reference)</td>
<td>1.0 (reference)</td>
<td>1.0 (reference)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient disposition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scene</td>
<td>4.67 (1.10–19.79)*</td>
<td>8.33 (1.62–42.96)*</td>
<td>7.33 (1.30–41.46)</td>
<td>6.29 (1.15–34.43)*</td>
<td></td>
</tr>
<tr>
<td>Transfer</td>
<td>1.0 (reference)</td>
<td>1.0 (reference)</td>
<td>1.0 (reference)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serum albumin percent change</td>
<td>2.06 (1.29–3.28)*</td>
<td>1.69 (1.02–2.80)</td>
<td>1.86 (1.05–3.31)</td>
<td>2.0 (1.02–1.41)*</td>
<td></td>
</tr>
<tr>
<td>Outcome LOS</td>
<td>1.05 (1.01–1.09)*</td>
<td>0.98 (0.93–1.02)</td>
<td>0.89 (0.84–0.95)</td>
<td>0.89 (0.83–0.95)*</td>
<td></td>
</tr>
<tr>
<td>Used immunosuppressive</td>
<td>2.22 (0.83–5.95)</td>
<td>2.55 (1.84–7.8)</td>
<td>5.40 (1.22–23.96)</td>
<td>4.56 (1.06–19.64)*</td>
<td></td>
</tr>
<tr>
<td>Had microbial resistance</td>
<td>4.04 (1.38–11.86)*</td>
<td>5.51 (1.58–19.24)</td>
<td>8.17 (1.55–43.0)</td>
<td>9.25 (1.84–46.42)*</td>
<td></td>
</tr>
<tr>
<td>Had other pre-existing infection(s)</td>
<td>4.27 (2.34–7.79)*</td>
<td>4.44 (2.16–9.12)*</td>
<td>5.46 (2.23–13.34)</td>
<td>5.46 (2.28–13.06)*</td>
<td></td>
</tr>
<tr>
<td>ISS (per 10 units change)</td>
<td>1.49 (1.18–1.89)*</td>
<td>1.68 (1.75–2.21)*</td>
<td>1.22 (0.90–1.66)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. OR = odds ratio; CI = confidence interval; BSI = bloodstream infections; CVC = central venous catheters; LOS = length of stay; ISS = Injury Severity Score.  
*bModel 2: Hosmer-Lemeshow goodness of fit $\chi^2 = 8.64, p = .374$, Cox & Snell $R^2 = .315$, Nagelkerke $R^2 = .549$.  
*Significant 95% CI.
specificity refers to the proportion of patients who did not have nosocomial BSI and were correctly classified as such. Sensitivity refers to the proportion of patients who had nosocomial BSI and were correctly identified as such by the model. The model’s positive predictive value refers to the probability that it correctly predicts that a patient will develop nosocomial BSI. The negative predictive value of the model refers to the probability that it correctly predicts that a patient will not develop nosocomial BSI. Although all classification indices provide important information, clinicians are often more interested in the predictive values of a model because predictive values provide information about the ability of the model to predict an outcome prior to its actual occurrence.

In many instances, classification indices are inversely related such that sensitivity decreases as specificity increases, and vice versa. This requires that clinicians make an informed decision as to which classification index is most important in a given situation. When studying an outcome variable that entails serious consequences such as nosocomial BSI, the clinician may decide to pay special attention to the negative predictive value of the model to minimize the chance of incorrectly concluding that a patient is at low risk of developing nosocomial BSI. Our data show that Model 3, which excluded the ISS and included the SMIS, had the best negative predictive value (93%), sensitivity (60%), and positive predictive value (76.4%). Model 3 also had a specificity of 97.4%, which was comparable to that of Model 1 (98.6%). Thus, Model 3 had the best overall predictive value as a screening tool.

The ISS has little clinical significance beyond its ability to predict the development of nosocomial BSI. In contrast, clinicians with knowledge of the SMIS can use it to minimize the risk of nosocomial BSI. Clinicians are urged to be vigilant in observing infection control practices with regard to preventing contamination of invasive device. For example, although clinicians may not be able to control the number of blood units a patient needs, more aggressive precautions against contamination and infection among patients with massive blood transfusion are warranted. In addition, knowledge of the impact of the SMIS on the risk of developing nosocomial BSI suggests a need for policy development and future research. Given the knowledge that the risk of nosocomial BSI increases as the number of central venous catheters inserted increases, nurses and other clinicians can collaborate to develop policies that discourage routine changes of these catheters as per the guidelines developed by the CDC (2002). At present, we do not know specifically how the SMIS contribute to the risk of developing nosocomial BSI. Therefore, we recommend that further research concerning the SMIS be conducted to identify the unique factors concerning these surrogate markers that contribute to nosocomial BSI. Knowledge of these factors may set the stage for future infection control guidelines and intervention studies aimed at preventing nosocomial BSI.

The ISS is one of a number of available anatomic, physiologic, and combined injury severity indices. It is, therefore, difficult to ascertain whether our findings would be replicated if other injury severity indices were used. This study demonstrates the clinical and theoretical importance of surrogate markers in research, but clinical implications need to be interpreted within the context of the observational cohort design, in which definitive causality cannot be inferred.

### Table 4. Trend of Change in the Classification Indices Across the Three Examined Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>18.2</td>
<td>98.6</td>
<td>71.4</td>
<td>87.0</td>
</tr>
<tr>
<td>Model 2</td>
<td>52.7</td>
<td>96.4</td>
<td>72.5</td>
<td>91.9</td>
</tr>
<tr>
<td>Model 3</td>
<td>60.0</td>
<td>97.4</td>
<td>76.7</td>
<td>93.0</td>
</tr>
</tbody>
</table>

Note. PPV = positive predictive value; NPV = negative predictive value. Values are in percentages.

Accepted for publication March 24, 2005.

The views expressed in this article do not necessarily represent those of the US Food and Drug Administration.

Corresponding author: Maher M. El-Masri, PhD, RN, University of Windsor, Faculty of Nursing, 401 Sunset, Health Education Center Room 128, Windsor, Ontario, Canada (e-mail: melmasri@uwindsor.ca).

### References


Background: Cost-effectiveness analyses are increasingly recommended to evaluate the effectiveness of health interventions. Determining the costs associated with delivery of a particular intervention is essential in conducting a cost-effectiveness analysis. Yet, there are few guidelines available to assist investigators in how to assess intervention costs associated with the personnel portion of an intervention.

Objectives: To describe the use of time studies in calculating the program costs of personnel for use in future cost-effectiveness analysis of health interventions.

Methods: The literature on calculating intervention costs for use in cost-effectiveness analyses is reviewed. The process for conducting a time study for determining personnel costs in delivering an intervention and a step-by-step example from a time study are used to illustrate how personnel costs associated with delivery of the intervention can be separated from those costs associated with implementation of research procedures in the determination of research costs.

Conclusions: Time studies provide a good estimate of part of the cost of implementing an intervention that is often difficult to determine—personnel time. The design of the time study should consider intervention components, staff involvement, and the time period for data collection.

Key Words: cost analysis · intervention · time studies
available from the health economics literature.

Determining program costs for personnel time can be complicated. Time spent on program implementation should be separated from time spent on tasks (e.g., the completion of a research instrument that would not be administered in a clinical setting or the obtaining of informed consent to participate in research) solely for research purposes. Examples of methods used to determine workers’ time include work sampling and time-motion studies. Work sampling involves using a sample of workers’ time and extrapolating the results to the total time spent in various activities. Time-motion studies (or time studies) involve direct determination of the time spent in those activities, through either direct observation or the workers’ self-report. While these methods have been used previously to determine productivity or to address staffing patterns, this article addresses the use of these methods for determining the cost of an intervention.

An example of work sampling that separates out research costs for a nursing intervention was described by Oddone, Weinberger, Hurder, Henderson, and Simel (1995). This intervention involved inpatient and outpatient counseling of patients who had been diagnosed with diabetes mellitus, congestive heart failure, or chronic obstructive pulmonary disease. The primary aim of this intervention was to reduce hospital readmissions. Patients were randomized to either the treatment group, which received intensive follow-up by nurses, or the usual care control group. In this study, the time spent in care activities was estimated by providing nurses with randomized beepers. When the beeper sounded, nurses would record their activity. The percentage of counts they recorded in each activity was extrapolated to estimate the percentage of total time spent in that activity. One advantage of this type of study is that the research costs can be removed easily from the intervention cost estimates. Research costs were substantial in this study, accounting for 42.5% of the total time spent working. However, no validation was conducted to determine whether nurses accurately reported the activity they said they were engaged in, or whether the proportion of times they recorded an activity was an accurate estimate of the proportion of total time spent in those activities.

The recommendations of the panel on cost-effectiveness are to include costs of a health worker’s time in the numerator of a CEA, and to evaluate this time by the wage rate in the labor force (Weinstein et al., 1996). A time study is the most direct way to calculate the amount of time that is devoted to an intervention and also to be able to separate the cost of research.

Time Study Exemplar

The following is an example of a time study done for a fall prevention intervention. The Fall Evaluation and Prevention Program (FEPP), funded by a grant from the National Institute of Nursing Research and the Office of Research on Women’s Health, National Institutes of Health (R01 NR05107), was designed to test a nurse-led, home-based program for the prevention of falls in community dwelling, high-risk older women. Participants were randomized to two groups: (a) an intervention group receiving the fall prevention program, consisting of exercise, education, and tailored risk reduction counseling; and (b) a control group receiving general health education. Both groups received a 12-week intervention involving biweekly home visits alternating with biweekly telephone calls, followed by a 16-week tapered computerized telephone intervention.

Other studies addressing the cost-effectiveness of implementing fall prevention programs have been conducted (Robertson, Devlin, Gardner, & Campbell, 2001; Robertson, Devlin, Scuffham, et al., 2001; Robertson, Gardner, Devlin, McGee, & Campbell, 2001; Rizzo, Baker, McAvay, & Tinetti, 1996), and while these have separated research costs, they have not included details of how this was done. In addition, while these studies included costs such as personnel time, supplies, and overhead, some have also included airfare and overnight accommodations for personnel. While these costs may be appropriate to include in a CEA, they potentially inflate program cost estimates for others who are considering implementing the intervention. The following example provides a more detailed description of how research-related expenses can be determined and separated from the program costs.

For the FEPP study, five categories of intervention costs were analyzed for their potential applicability to the time study. These included Development and training costs such as expenses associated with the development of the intervention protocols, including production of educational materials and the algorithm for the computerized telephone system, preparing the operational manual, and training research staff in the assessment and intervention procedures. Enrollment costs included staff time required to screen potential participants following initial contact to identify those who met eligibility criteria and were willing to participate. Overhead costs included estimates of office space rental, utilities, and supplies associated with the program. Equipment costs included health assessment equipment, participant educational materials, exercise equipment, and other items necessary to complete the assessment and intervention protocol. Staff-related expenses included the mean nurse contact time; mean time required to complete documentation of visits; mean time spent in round-trip travel to participants’ homes; mean mileage costs associated with all assessment and intervention visits; mean time needed for consultation among the nurse, research physician, and participant’s physician; and mean time required for telephone calls to participants. These costs were based on documentation collected over a 2-week period during the middle part of the study. The time study focused primarily on the enrollment costs and the staff-related expenses.

Time Study Procedures

In a time study, the first step is identifying the general categories of costs. The three major steps in identifying costs are to (a) identify the relevant cost items, (b) measure the amount of resources used, and (c) place a value on those resources (Brouwer, Rutten, & Koopmanschap, 2001). It is important
at this step to get input from all the individuals involved in the intervention, particularly those directly delivering the intervention. Examples of time assessed from the FEPP study included travel time and mileage to and from participants’ homes, enrollment time, length of assessment visits and intervention visits, and other miscellaneous administrative time, such as consultations with participants’ physicians, photocopying materials, voice mail management, electronic mail management, staff meetings to discuss issues that arise, and “downtime in the field” as nurses waited between scheduled home visits. The largest allocations of administrative time in the FEPP study were the consultations with participants’ physicians to determine whether the person could participate in the exercise program, voice mail management, meetings, and downtime in the field.

The next step in a time study is to create a general outline for the
time chart or data-collection form (Figure 1). It is necessary to include all the times a person is likely to work, especially if employees may be working from home or work after hours. Generally, it is effective to allocate one sheet of paper to a single day and to break down the time periods into 15-min increments. All possible activity codes should be included and coded. If research procedures are embedded within the intervention, this activity should be separated from those activity

<table>
<thead>
<tr>
<th>Mileage</th>
<th>Time</th>
<th>Activity code</th>
<th>Comments</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Enrollment costs</th>
<th>Time</th>
<th>Activity code</th>
<th>Comments</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Staff-related expenses</th>
<th>Time</th>
<th>Activity code</th>
<th>Comments</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Miscellaneous administration</th>
<th>Time</th>
<th>Activity code</th>
<th>Comments</th>
</tr>
</thead>
</table>

FIGURE 1. Example of the documentation form used for FEPP time study. R indicates time spent on research-related activities. Copyright 2002 by the Center for Gerontological Nursing, University of Minnesota, Minneapolis.
codes that are specific only to the intervention. For example, in the FEPP study, the research nurses administered several instruments measuring exercise constructs at the first and last home visit in the experimental group only. These were used only as a test of the theoretical model guiding the study and were not part of the intervention itself. Another major component was derived from the intention to compare the number of falls in the intervention group against the number of falls in the control group. This control group consisted of women who received information on general health education only. While this health education group was necessary for the research portion of the study, it would not be utilized if the intervention were to be applied more broadly. These data were collected for the time study and for future research planning purposes, but they were excluded from analyses that assess the cost of the fall prevention intervention because they were purely research-related. It is necessary also to include a space on the form for general comments about each time category so that miscellaneous information can be added by personnel as needed. For example, in the FEPP study, one nurse spent an unexpected amount of time commuting to a participant’s home, but was able to explain that there had been a major traffic accident on the freeway at that time.

Once a data-collection instrument for a time study has been developed and approved by all of those working on the project, a pilot test of the instrument is necessary. This allows those who will actually be using the instrument to assess whether it applies to their experiences or if new codes and categories need to be added. It is also recommended that the pilot test be conducted using more than one individual as variability may exist in how the intervention is implemented and how codes are interpreted. Pilot testing on all types of visits or intervention components, so that all categories are utilized, will help determine if these categories need to be revised. Once the pilot study is complete, then the data-collection instrument should be revised accordingly. The more descriptive the categories are, the better will be the capture of data. Writing operational definitions of the categories or short descriptions of the type of activities covered in a particular category is essential. General categories like “other” or “miscellaneous” may be necessary, but if a large amount of time is put into these “other” categories, the data will not be as useful.

Once the instrument is revised, the time study can be implemented. It is important to select a time frame that is representative of the overall study from which to estimate costs. In the FEPP study, the intervention involved a series of home visits and telephone calls. These visits consisted of enrollment visits (to assess the person’s eligibility and fall risk status) and intervention visits (where the intervention was administered). In addition, each intervention visit presented different material, so the length of the visits may have varied. If different types of visits are included in an intervention, it is important to choose a time period for the time study that will allow the assessment of the cost of each phase. In addition, some points in time may not adequately represent the cost of the intervention. The significance of learning curves on cost estimates is emphasized by Brouwer et al. (2001). Choosing a time period that is too early in the intervention may overestimate the time necessary if nurses are still learning the study protocol. Likewise, choosing a time period that is too late may result in missing key components of an intervention if one type of visit on all participants is complete. If individuals need to be assessed prior to the intervention, then choosing a time period, after all assessments have been completed, will limit the generalizability of the cost estimate to include this type of visit. In the FEPP study, a time period of 2 weeks that was midway through the intervention period was selected.

Once the data are collected, they can be entered into a spreadsheet for analysis. The totals for each code or type of activity can be summed and averaged, for an average amount of time per activity. As stated above, if research activities are present, these should be separated from the nonresearch activities. Once the time estimates are complete, then dollar amounts in wages can be applied for each time period to arrive at a cost estimate. This allows for differences in the job classification of the person completing each task and geographic variations in salary. Assumptions regarding the time study should be clearly stated; for example, the volume of visits would be constant across participants or the amount of downtime would be constant. For a study involving home visits, a consistent mileage cost should be used, such as that allowed by the federal government for tax purposes.

Time studies provide a good estimate of the personnel cost of implementing an intervention for use in cost-effectiveness analyses. The design of the time study should consider the intervention components, related activities, staff involvement, and the time period for data collection. Information obtained can be used also to redesign interventions for future testing.

References


Accepted for publication March 1, 2005.

Funded by the National Institute of Nursing Research and the Office of Research on Women’s Health, National Institutes of Health (R01 RO05107). The authors thank Lois Gildea, Melinda Monigold, Carrie Gomez, Kris Talley, and Jennifer Peters for their contributions to the time study and Sean Lamb-Vosen for data entry.

Corresponding author: Mary J. Findorff, PhD, RN, School of Nursing, University of Minnesota, 1-160 WDH, 308 Harvard St SE, Minneapolis, MN 55455 (e-mail: find0003@umn.edu).


