CJNR 2004, Vol. 36 Nº 3, 3–5

EDITORIAL

Nursing's Ethos of Caring and Its Support for a Single-Tiered Health-Care System

Canadians' commitment to a publicly funded universal health-care system was re-affirmed when they went to the polls this past June and re-elected a government that they believed would preserve the current system and prevent the creation of a two-tiered system: one for the rich and the one for the rest of society. Despite widespread discontentment with the current system, Canadians voted in a minority Liberal government, sending a powerful message to Ottawa that they wanted their "broken" but beloved health-care system fixed. All was not right, but they were not ready to give up — not just yet.

Since the introduction of medicare almost 40 years ago, there have been many attempts to erode its principles, among which is equal access for all. While acknowledging the weaknesses in the system, nurses and nursing organizations have consistently advocated for and supported medicare. Nurses have never wavered in their support for the principles of the *Canada Health Act* despite the incredibly harsh conditions under which they have laboured.

The past decade has been brutal to nursing. The system under which nurses work has not been as generous, supportive, committed, and loyal to them as nurses have been to it. It is well documented that nurses have shouldered a disproportionate share of the burden wrought by financial cuts, downsizing, and mergers. Nurses have been marginalized, de-professionalized, and demoralized. They have endured abuses and working conditions that few other health professionals have had to face. They have paid dearly with their own health, frozen and lost wages, elimination of jobs, a decimated leadership structure, working conditions that border on the inhumane, loss of status, workplace violence and abuse, shortages, recruitment and retention difficulties — the list goes on. And yet despite the deplorable working conditions under which they care for patients and their families, nurses have remained steadfast in their support of a single-tiered, nationally funded health-care system. The question is why. Are nurses masochists? Angels? Paralyzed? Why have nursing organizations not advocated for a return to privately funded health care?

The answer may be found in nursing's ethos of caring. Caring has been nursing's banner, and, for many, caring and nursing are synonymous. In 1970 the slogan Nurses Care; Physicians Cure was created to distinguish nursing from medicine. This was a time when nursing was seeking its own identity. It was an unfortunate slogan inasmuch as it sent the erroneous message that only nurses care. The reality is that the majority of those who choose a career in one of the many helping professions do so because they are dedicated and committed to people in need. They want to contribute to the betterment of humanity. Nurses care. Doctors care. Other health professionals care. However, there are many models of caring and many ways of expressing caring. The different models of caring may explain why one health-care profession supports a single-tiered system while another supports a two-tiered system.

During my sabbatical year in Berkeley in the aftermath of 9/11, I was fortunate to have an opportunity to meet with Dr. George Lakoff, professor of linguistics and cognitive science at the University of California, Berkeley, to discuss a number of issues that had been troubling me about nursing, nursing's role in the health-care system, and the nurse-doctor relationship. Dr. Lakoff referred me to his book *Moral Politics* (1996). The book advances a new framework to explain why liberals and conservatives approach a problem very differently. This framework proved enlightening; it provided me with some insights into the nature of nursing's form of caring and helped to explain why nurses continue to be committed to a single-tiered health-care system despite the battering they are taking.

Drawing on the analogy of different family structures and ways of functioning, Dr. Lakoff describes two basic family forms to represent two approaches to moral and political action: the Nurturant Parent Model and the Strict Father Model. The two models, based on different worldviews, give rise to different moral systems and different modes of reasoning and discourse, and lead to very different ways of acting. The Nurturant Parent Model stresses social responsibility, social and individual ends, and individual rights and freedoms, whereas the Strict Father Model stresses survival of the fittest, taking responsibility for oneself, self-reliance, and individual rights over social responsibility.

Nurses generally subscribe to the Nurturant Parent Model, because nurses bear witness to suffering. Nurses are privy to the most intimate aspects of a person's and a family's life. Nurses know the hardships endured during illness and how these hardships lead to increased vulnerability when access to affordable care is limited. They know how the vulnerable become more vulnerable, the needy more needy, the despondent more despondent when health care is not fully accessible or affordable. Nurses often put society's interests above their own.

Nurses know that the health of the nation is put in jeopardy when the health-care system is compromised and becomes dysfunctional. During the past decade nurses have sounded the alarm, pointing out that the system is not working. Eventually the cracks became so wide, the fissures so deep, and nurses' voices so loud that politicians and policy-makers had to take heed. Money has been committed to finding ways to address the nursing shortage. Some recognize that poor working conditions seriously compromise patient care. Others have yet to make this connection despite research evidence that directly links nurses' staffing ratios to proper care (Aiken et al., 2002). More of this type of research is needed. Nurse researchers, clinicians, educators, and administrators must continue to record their own experiences and those of their patients and communicate these findings to others. There is a danger that when the crisis lifts, politicians, policy-makers, and health-care administrators will turn their attention elsewhere and nurses and nursing will be ignored. Nurses need to stand guard and continue to fight, armed with evidence, until the health-care system is fundamentally changed and nurses achieve the power, status, and working conditions they need in order to care for patients and families in a manner that is consistent with their ethos of caring.

Within the health-care milieu, nurses are among the strongest advocates of retaining the single-tiered system. However, if the government fails to properly support nursing and to radically reform nurses' working environment, it will find that its most loyal ally has deserted it. Even the Nurturant Parent cannot support a morally bankrupt environment. Without the support of nursing, the demise of the single-tiered system is inevitable. And it is more than the health-care system that will be lost. Canadians' sense of identity and this country's moral compass are also in jeopardy.

Laurie N. Gottlieb
Editor-in-Chief

References

Aiken, L. H., Clarke, S. P., Sloane, D. M., Sochalski, J., & Silber, J. H. (2002). Effects of nurse staffing on nurse burnout and job dissatisfaction, and patient deaths. *Journal of the American Medical Association, 288,* 1987–1193.

Lakoff, G. (1996). *Moral politics: What conservatives know that liberals don't.* Chicago: University of Chicago Press.

CJNR 2004, Vol. 36 Nº 3, 7–18

Guest Editorial and Discourse

Rethinking the Research-Practice Gap: Relevance of the RCT to Practice

Souraya Sidani

The research-practice gap is not a new phenomenon in nursing. It was first observed about 50 years ago. The results of surveys conducted in various countries and at different points in time have consistently indicated that a small percentage of nurses use research-based evidence to guide their practice (e.g., Brett, 1987; Coyle & Sokop, 1990; Ketefian, 1975; Kirchhoff, 1982; Parahoo, Barr, & McCaughan, 2000; Rodgers, 2000). Attention to the research-practice gap has resurfaced at a time when evidence-based practice is viewed as a means for enhancing the quality of health and nursing care. The implementation of evidence-based nursing practice requires that nurses critically review the research evidence supporting the effectiveness of interventions, synthesize it, and translate it into practice guidelines. Yet, anecdotal and empirical observation suggests that the updating and use of research findings in everyday practice is rather limited (Naylor, 2003).

Barriers to research utilization have been identified in four areas: the social system in which nurses are employed, nurses' research values and skills, quality of research, and methods for communicating research (Funk, Tornquist, & Champagne, 1995). Various strategies have been proposed to address these barriers and subsequently the research-practice gap. For instance, unit-based journal clubs and workshops focusing on the critical appraisal of research have been established in an effort to increase nurses' research knowledge and skills. Other organizational strategies to support the research-utilization endeavours of nurses include the following: pairing of researchers and clinicians to work on developing practice guidelines, making computers available in the practice setting to facilitate access to research literature, and using research or evidence-based practice as a key value (Estabrooks, 2003; Funk et al.; LeMay, Mulhall, & Alexander, 1998). Another strategy is to improve the methods for disseminating research findings to practitioners (Oxman, Thomson, & Davis, 1995).

Despite our understanding of the factors that lead to the research-practice gap and our exploration of strategies to promote evidence-based practice, only to a limited extent are research findings being used to guide practice. Therefore we must re-think or re-prioritize the factors that contribute to the research-practice gap and adopt strategies to close it. Specifically, the perceived irrelevance of research to practice has received little attention as a barrier to research utilization. Research methods are far removed from the reality of practice and therefore generate findings that have limited applicability. The randomized controlled trial (RCT) is a case in point. It is crucial that we modify available research methods and/or design alternative ones so that they are in line with practice and produce clinically relevant findings. The changes in research methods may facilitate the updating and application of clinically relevant findings, and thereby help to close the research-practice gap and promote evidence-based practice.

The Roots of the Research-Practice Gap

In recent reports of quantitative and qualitative studies and clinical observations, two specific barriers have been identified as hampering research utilization in different clinical settings and different countries: unclear implications for practice, and lack of generalizability and relevance to the practice setting (e.g., LeMay et al., 1998; Nilsson Kajermo, Nordstrom, Krusebrant, & Bjorvell, 2000; Retsas, 2000). This concern on the part of nurses about the limited generalizability and clinical relevance of research findings is echoed by scholars in applied disciplines including medicine, nursing, and the behavioural sciences. Scholars question the ability of available research approaches and methods to produce clinically meaningful results and the applicability of research findings to everyday practice (e.g., Clarke, 1995; Ferguson, 2004; Gross & Fogg, 2001; Lindsay, 2004; Pincus, 2002; Sidani & Epstein, 2003; TenHave, Coyne, Salzer, & Katz, 2003).

The discrepancy between the way that research is conducted and the way that practice is conducted, and between the knowledge that is generated through research and the knowledge that is needed to guide practice, is a key factor in the research-practice gap. Research approaches and methods are developed in the basic sciences. They are adopted and applied without critical analysis of their suitability and utility to the study of phenomena in applied disciplines such as nursing. Qualitative approaches to research are grounded in the social sciences, quantitative approaches in the physical sciences. The application of the methods used in these approaches may not be fully consistent with the study of health-related phenomena of interest in nursing, and with the need to form a

comprehensive understanding of the patients' presenting problem, preferences, and responses to care as they occur in practice. For example, the RCT, considered the standard in intervention evaluation, is far removed from practice (Sidani, Epstein, & Moritz, 2003). It was developed to examine the impact of innovations in the field of agriculture. Its use in investigating the effectiveness of nursing interventions that are delivered to human beings in order to manage their often complex health-related problems is questionable. The RCT design is based on a perspective of science and assumptions about human beings and their response to treatment that are incongruent with those embraced by nursing. Further, the characteristic features of the RCT do not reflect the reality of practice (Sidani et al.), particularly with the current emphasis on patient-centred care. The assumptions underlying the RCT design are contradictory to those underpinning the patient-centred approach to care, in which each patient is considered a unique person, having individual needs and preferences, and responding differently to interventions. In patient-centred care, the selection and implementation of interventions are mindful of and responsive to the characteristics and preferences of individual patients (Lauver et al., 2002; McCormack, 2003).

Closing the Research-Practice Gap: Advances in Research Methods

Faced with the discrepancy between the knowledge generated by research and the knowledge needed to guide practice, and with the questionable ability of available research approaches and methods to produce clinically relevant and applicable findings, scholars in various applied disciplines, including nursing, have cited the need for alternative research methods (e.g., Gross & Fogg, 2001; Sidani et al., 2003; TenHave et al., 2003; Tunis, Stryer, & Clancy, 2003). Over the years there has been an increase in the number of articles citing the limitations of available research methods, suggesting these be modified or proposing innovative ones. The discourse on the discrepancy between research and practice has focused primarily on the RCT design used to evaluate the effectiveness of interventions. Delivering interventions is the essence of nursing. Examining the effectiveness of interventions, therefore, is critical to nursing practice. Effectiveness is concerned with the robustness of intervention effects when implemented by various nurses, with different patient populations, in the real world of nursing practice (Whittemore & Grey, 2002).

Questions have been raised about three aspects of the RCT design with regard to its utility for evaluating the effectiveness of interventions and for generating clinically significant findings. These are: overall design,

assignment of participants to study groups, and analysis of data. For each aspect, the discrepancy between research and practice and ways to minimize it are discussed.

Overall Design

The characteristic features of the RCT impose artificial boundaries that limit the relevance and generalizability of the findings. The first feature is the careful selection of participants based on a set of rigorous inclusion and exclusion criteria. These criteria are a means for controlling confounding variables, which increases the likelihood that significant effects will be detected and increases confidence in attributing the effects to the intervention. Yet this careful selection limits the number of eligible participants and confines the sample to certain subgroups of the target population. The end result is lack of generalizability to the full range of subgroups that make up the population of patients encountered in everyday practice (Brown, 2002; Pincus, 2002; Sidani et al., 2003; TenHave et al., 2003; Tunis et al., 2003). The second feature, random assignment of participants to study groups, is discussed in the next section. The third feature is manipulation of implementation. The intervention is administered to one group of participants and withheld from the other; it is delivered in a standardized way, at a fixed dose. Standardized implementation at a fixed dose ensures consistency of exposure and response to the treatment, which in turn enhances the statistical power to detect significant intervention effects (Lipsey, 1990; Shadish, Cook, & Campbell, 2002). This RCT feature is not consistent with the individualized nursing interventions that make up the patient-centred approach to care. Individualization involves the adaptation of some elements and/or dose of the intervention in order to respond flexibly to patients' needs and preferences (Lindsay, 2004; Sidani et al; TenHave et al.). Furthermore, the results of an RCT in which the dose of the intervention administered is not accounted for in the analysis do not provide an accurate estimate of effects, nor do they indicate the optimal dose needed to produce the desired outcomes (Pincus). The fourth feature of the RCT is the control of factors in the setting in which the intervention is implemented. These factors may interfere with the delivery of the intervention and/or the achievement of outcomes. They are controlled experimentally by maintaining them constant across all instances of intervention delivery. Such control is unfeasible in a complex healthcare system. RCT results may not be applicable across a range of settings (Conrad & Conrad, 1994).

The current trend is to conduct a pragmatic or practical clinical trial to evaluate the effectiveness of an intervention (Tunis et al., 2003). In this type of trial, participants with diverse characteristics or backgrounds are

recruited from various settings in order to ensure the representation of all subgroups that make up the population of patients seen in practice. Subgroup analysis is recommended as a way of determining the extent to which each subgroup benefits from the intervention (Brown, 2002; Sidani et al., 2003; Tunis et al.). The interventions selected for comparison in a clinical trial should be clinically relevant; that is, they should be feasible in practice and allow for flexibility in implementation and dose (Concato & Horwitz, 2004; TenHave et al., 2003). Careful monitoring of the fidelity with which the intervention is implemented (Judge Santacroce, Maccarelli, & Grey, 2004) and of the dose administered is essential. Dose-response analysis provides information about the optimal dose needed to produce the desired outcomes (Sidani, 1998). Identifying and measuring the characteristics of the setting that could influence effectiveness, and examining their impact on the outcomes, is a useful means of determining aspects of the practice setting that should be modified in order to enhance implementation and effectiveness (Sidani & Braden, 1998).

Assignment of Participants

Random assignment to experimental and control groups serves to generate a balanced distribution, between the groups, of participants with similar characteristics. This initial group equivalence provides a means for controlling the influence of extraneous factors on the outcomes, and therefore for enhancing the validity of the claim that it is the intervention that caused the outcomes (Shadish et al., 2002). Random assignment has two limitations. The first relates to the notion that the initial group equivalence achieved by random assignment is probabilistic, implying that random assignment increases the likelihood, but does not guarantee, that the study groups are exactly comparable on all characteristics that may affect the outcomes (Rossi, Freeman, & Lipsey, 1999; Vandenbroucke, 2004). Because of the probabilistic nature of initial equivalence, the groups' mean scores may differ on baseline variables; however, such differences are considered to be due to chance, even if statistically significant (Rossi et al.). No matter how small they are, such between-group differences cannot be ignored, as they may be clinically meaningful and are associated with post-test differences in the outcomes, as indicated by the results of meta-analytic studies (Heinsman & Shadish, 1996; Sidani, in press).

The second limitation relates to the fact that participants have treatment preferences, which, when ignored in random assignment, may pose threats to validity. In the consent process, participants are informed about the experimental and control conditions of the investigation. They may perceive the treatment options as unequally attractive and may have a

preference. Participants who do not receive their preferred treatment may become dissatisfied with the assigned treatment and fail to comply with the intervention and/or drop out of the study. The result is a biased estimate of the intervention effects (Bottomley, 1997; Torgerson, Klaber-Moffett, & Russell, 1996). Therefore, the observed intervention effects may not be reproduced when the intervention is administered to patients in a way that is consistent with their preferences, as in the patient-centred approach to care.

Recent methodological advances include modelling of potentially confounding factors as well as participant treatment preferences. Researchers can identify participant characteristics that are conceptualized to influence the outcomes, measure them reliably, and adjust for their influence statistically, even if the participants were randomly allocated (Vandenbroucke, 2004). Alternatively, they can examine the direct and/or indirect effects on the outcomes to identify the participants who benefited most from the intervention (Sidani et al., 2003). Taking participants' treatment preferences into account when allocating them to the study groups or when conducting the statistical analyses is the most recent advance in the design of effectiveness studies. Specifically, the partial RCT is a means for bringing research designs more in line with practice and for enhancing the clinical relevance of research findings (Sidani et al.; TenHave et al., 2003). It also holds promise for minimizing the threats associated with random assignment (Corrigan & Salzer, 2003; McPherson & Britton, 2001).

Data Analysis

The data obtained in intervention evaluation studies are analyzed using statistical tests, which compare the mean outcome scores of the experimental and control groups. Statistically significant differences in the post-test outcome mean scores provide the evidence to infer that the intervention is effective. The model underlying these statistical tests (i.e., *t* test or *F* ratio) is based on the assumption that the intervention effects show up as a change, of a constant value, in the participants' post-test outcome scores. The constant change in the scores of participants who received the intervention creates the difference between the experimental and control groups expected in the outcomes measured at post-test (Bryk & Raudenbush, 1988; Lipsey, 1990). Individual differences in outcome achievement are possible; however, they are represented in the within-group variance, which is considered error variance. A large within-group variance reduces the statistical power to detect significant intervention effects (Lipsey). Yet, a large within-group variance indicates that the participants have responded to the intervention to different degrees (Bryk & Raudenbush). Information about individual or subgroup differences

in response to treatment is of the utmost importance to clinicians (Jacobson & Truax, 1991). Researchers are strongly encouraged to supplement the traditional group-level analysis with analyses at the subgroup (Brown, 2002) or individual level using advanced statistical techniques such as Hierarchical Linear Models (HLM; Raudenbush & Bryk, 2001).

Reliance on tests of statistical significance to demonstrate the positive impact of an intervention is potentially misleading (Basch & Gold, 1986). Statistically significant findings do not address the size, strength, or clinical significance of intervention effects (Rothstein & Crabtree Tonges, 2000). Statistical significance indicates that the observed effects are real, reliable, and unlikely to be due to chance. Clinical significance "refers to the benefits derived from treatment, its potency, its impact on clients, or its ability to make a difference in people's lives" (Jacobson & Truax, 1991, p. 12). It should be assessed in order to validate the utility of the intervention in addressing the clinical problem of interest. LeFort (1993) provides a comprehensive overview of the various approaches for examining clinical significance. While the statistical approach (i.e., calculating the effect size) is the most common, individual approaches are gaining momentum. Individual approaches indicate the proportion of participants whose outcome improved. They therefore complement the results of subgroup- or individual-level analyses and are consistent with the patient-centred approach to care. The results of individual approaches present the findings in terms that are familiar to and consequently easily understood by clinicians. Normative approaches have great potential but require the availability of cut-off scores or normative values for the instruments measuring the outcomes of interest to nursing. Assessment and reporting of the clinical meaningfulness of results obtained in intervention effectiveness studies is key to reducing the research-practice gap.

An Invitation

Despite all the efforts to promote evidence-based practice, research findings are still not being widely used to guide practice. The current trend is to attribute the research-practice gap to the lack of generalizability of research findings to the real world of practice, as well as to the lack of clinical relevance of research results. We will have to modify available research methods and/or design alternative ones that are more in line with practice and that produce more clinically significant results. The literature addressing this issue has increased over the years. Journals are now devoting entire issues to topics such as methodological problems in clinical research or the exploration of recent developments in research methods. An invitation has been extended to researchers to move towards an alternative paradigm for the evaluation of intervention effectiveness

(Gross & Fogg, 2001). In fact, the move to an alternative paradigm has taken place (Concato & Horwitz, 2004; Sidani et al., 2003).

This issue of *CJNR* extends the discourse on research methods in two ways. First, it invites the reader to critically appraise the utility of research methods in terms of the congruence between the assumptions that underlie them and the assumptions that underlie nursing practice, and in terms of their ability to produce clinically relevant results. Second, it encourages the reader to thoughtfully design research methods and test their ability to generate reliable, valid, and clinically meaningful results that will provide a knowledge base to guide current practice, where the emphasis is on patient-centred care. The papers in this issue of the Journal present advances in research approaches and methods that expand on those suggested under the alternative paradigm and that hold promise for the development of a sound and clinically useful knowledge base.

In two papers, the authors advance modifications to the current approach to research that are consistent with the conceptualization and operationalization of nursing practice. Redman and Lynn review the inconsistencies between research designs and methods and between the conceptualization and implementation of patient-centred care. They argue that the demand for standardized interventions and measures, for the careful selection and random assignment of participants, and for controlling or ignoring extraneous and contextual factors is incongruent with the need for interventions that are tailored to patients' needs and preferences. Redman and Lynn recommend knowledge-development approaches that pertain to the effectiveness of patient-centred, tailored interventions; the most challenging of such approaches is the generation of measures that take account of individual values and preferences. Guruge and Khanlou reconceptualize the complexity and interrelatedness of multilevel social influences on health and well-being. They propose Participatory Action Research (PAR) as a useful means for identifying issues of importance to patients and the strategies patients need in order to promote their health. These essential features of PAR are consistent with those of patient-centred care, whose emphasis is serving patients' needs, values, and preferences.

The partial RCT is considered a modified design suitable for investigating patient-centred, individualized interventions. It permits the researcher to take participants' treatment preferences into account. Miranda reports that the majority of participants are unwilling to be randomly assigned and that their treatment preferences are influenced by several factors. These findings have implications for the design of studies evaluating the effectiveness of interventions and the design of materials and measures for eliciting patient preferences.

Marcellus encourages researchers to reconceptualize attrition from an ecological perspective in which factors related to the participant, the researcher, the study, and the environment affect the participant's decision whether to complete or drop out of a study. This reconceptualization guides the investigator in selecting strategies to minimize attrition and incorporating these into the study design. Marcellus recommends the active involvement of participants in the selection of strategies, which is consistent with the paradigm shift towards patient-centred research.

Lynn and McMillen present a methodology for constructing instruments that capture the perspective of patients and the variability in their responses to measures of satisfaction with care. The methodology comprises the use of (1) qualitative methods to elicit the patient's perspective and to generate items, and (2) the scale product technique to examine the importance of the items for each individual respondent. This methodology holds promise for constructing and using measures that take account of individual values and preferences, consistent with the patient-centred approach to research and care.

Several authors address the application of alternative statistical approaches to data analysis. Fox and colleagues demonstrate the utility of hierarchical linear models in analyzing the pattern of change in outcomes. The analysis is conducted at the individual level, which is consistent with the assumption underlying patient-centred care that participants vary in their responses to a treatment. Such findings complement traditional group-level results and are useful in delineating the clinical benefits of an intervention. Lucke provides an overview of the principles and techniques characterizing Bayesian statistics, and illustrates their use in systematic quantitative reviews of studies evaluating the effects of fall-prevention programs. The applicability of Bayesian statistics to intervention effectiveness research needs further exploration.

Snowdon presents the Personal Construct Theory and its accompanying methodology, the repertory grid technique, as a means for investigating complex, multidimensional phenomena of interest to nursing. The renewed interest in this theory and technique may pave the way for quantitative exploration of the meaning and significance of individual patients' personal belief systems.

The advances in research methods presented in this issue of the Journal represent the seed of a fruitful initiative to develop research approaches and methods that are consistent with the nursing perspective, that enable the study of individualized interventions reflective of patient-centred care, and that enhance the clinical relevance of findings. Concern about the ability of available research methods to produce clinically meaningful results has prompted the move towards an alternative research paradigm. We invite nurse researchers to re-think and improve on current

methodology in order to close the research-practice gap. Nurse researchers are in the best position to take the lead in this initiative. The legacy of their most prominent role model, Florence Nightingale, in carefully collecting and analyzing data, and using the results to guide practice, should be an inspiration to all.

References

Basch, C. E., & Gold, R. S. (1986). The dubious effects of type V errors in hypothesis testing on health education practice and theory. *Health Education Research, 1*(4), 299–305.

Bottomley, A. (1997). To randomise or not to randomise: Methodological pitfalls of the RCT design in psychosocial intervention studies. *European Journal of Cancer Care, 6,* 222–230.

Brett, J. L. (1987). Use of nursing practice research findings. *Nursing Research, 36,* 344–349.

Brown, S. J. (2002). Nursing intervention studies: A descriptive analysis of issues important to clinicians. *Research in Nursing and Health, 25,* 317–327.

Bryk, A. S., & Raudenbush, S. W. (1988). Heterogeneity of variance in experimental studies: A challenge to conventional interpretations. *Psychological Bulletin, 104*(3), 396–404.

Clarke, G. N. (1995). Improving the transition from basic efficacy research to effectiveness studies: Methodological issues and procedures. *Journal of Consulting and Clinical Psychology, 63,* 718–725.

Concato, J., & Horwitz, R. I. (2004). Beyond randomised versus observational studies. *Lancet, 363,* 1660–1661.

Conrad, K. J., & Conrad, K. M. (1994). Reassessing validity threats in experiments: Focus on construct validity. *New Directions for Program Evaluation, 63,* 3–25.

Corrigan, P. W., & Salzer, M. S. (2003). The conflict between random assignment and treatment preference: Implications for internal validity. *Evaluation and Program Planning, 26,* 109–121.

Coyle, L. A., & Sokop, A. G. (1990). Innovation adoption behavior among nurses. *Nursing Research, 39,* 176–180.

Estabrooks, C. A. (2003). Translating research into practice: Implications for organizations and administrators. *Canadian Journal of Nursing Research, 35*(3), 53–68.

Ferguson, L. (2004). External validity, generalizability, and knowledge utilization. *Journal of Nursing Scholarship, 36*(1), 16–22.

Funk, S. G., Tornquist, E. M., & Champagne, M. T. (1995). Barriers and facilitators of research utilization: An integrative review. *Nursing Clinics of North America, 30*(3), 395–407.

Gross, D., & Fogg, L. (2001). Clinical trials in the 21st century: The case for participant-centered research. *Research in Nursing and Health, 24,* 530–539.

Heinsman, D. T., & Shadish, W. R. (1996). Assignment methods in experimentation: When do nonrandomized experiments approximate the answers from randomized experiments? *Psychological Methods, 1,* 154–169.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12–19.

Judge Santacroce, S., Maccarelli, L. M., & Grey, M. (2004). Intervention fidelity. *Nursing Research, 53*(1), 63–66.

Ketefian, S. (1975). Application of selected nursing research findings into nursing practice: A pilot study. *Nursing Research, 24,* 89–92.

Kirchhoff, K. T. (1982). A diffusion survey of coronary precautions. *Nursing Research, 31,* 196–201.

Lauver, D. R., Ward, S. E., Heidrich, S. M., Keller, M. L., Bowers, B. J., Brennan, P. F., et al. (2002). Patient-centered interventions. *Research in Nursing and Health, 25,* 246–255.

LeFort, S. (1993). The statistical versus clinical significance debate. *Image: Journal of Nursing Scholarship, 25,* 57–62.

LeMay, A., Mulhall, A., & Alexander, C. (1998). Bridging the research-practice gap: Exploring the research cultures of practitioners and managers. *Journal of Advanced Nursing, 28*(2), 428–437.

Lindsay. B. (2004). Randomized controlled trials of socially complex nursing interventions: Creating bias and unreliability? *Journal of Advanced Nursing, 45*(1), 84–94.

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage.

McCormack, B. (2003). Researching nursing practice: Does patient-centredness matter? *Nursing Philosophy, 4,* 179–188.

McPherson, K., & Britton, A. (2001). Preferences and understanding their effects on health. *Quality in Health Care, 10*(Suppl 1), 161–166.

Naylor, M. D. (2003). Nursing intervention research and quality of care: Influencing the future of healthcare. *Nursing Research, 52*(6), 380–385.

Nilsson Kajermo, K., Nordstrom, G., Krusebrant, A., & Bjorvell, H. (2000). Perceptions of research utilization: Comparisons between health care professionals, nursing students and a reference group of nurse clinicians. *Journal of Advanced Nursing, 31*(1), 99–109.

Oxman, E., Thomson, M., & Davis, D. (1995). No magic bullets: A systematic review of 102 trials of interventions to improve professional practice. *Canadian Medical Association Journal, 153,* 1423–1431.

Parahoo, K., Barr, O., & McCaughan, E. (2000). Research utilization and attitudes towards research among learning disability nurses in Northern Ireland. *Journal of Advanced Nursing, 31*(3), 607–613.

Pincus, T. (2002). Limitations of randomized clinical trials in chronic diseases: Explanations and recommendations. *Advances in Mind-Body Medicine, 18*(2), 14–21.

Raudenbush, S., & Bryk, A. (2001). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Retsas, A. (2000). Barriers to using research evidence in nursing practice. *Journal of Advanced Nursing, 31*(3), 599–606.

Rodgers, S. E. (2000). The extent of nursing research utilization in general medical and surgical wards. *Journal of Advanced Nursing, 32*(1), 182–193.

Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (1999). *Evaluation: A systematic approach* (6th ed.). Thousand Oaks, CA: Sage.

Rothstein, H., & Crabtree Tonges, M. (2000). Beyond the significance test in administrative research and policy decisions. *Journal of Nursing Scholarship, 32*(1), 65–70.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental design for generalized causal inference.* Boston: Houghton-Mifflin.

Sidani, S. (1998). Measuring the intervention in effectiveness research. *Western Journal of Nursing Research, 20*(5), 621–635.

Sidani, S. (in press). Random assignment: A systematic review. In R. R. Bootzin & P. E. McKnight (Eds.), *Measurement, methodology, and evaluation: Festschrift in honor of Lee Sechrest*. Washington: APA Books.

Sidani, S., & Braden, C. J. (1998). *Evaluating nursing interventions: A theory-driven approach*. Thousand Oaks, CA: Sage.

Sidani, S., & Epstein, D. R. (2003). Clinical relevance of RCT findings. *International Nursing Perspectives, 3*(1), 49–56.

Sidani, S., Epstein, D. R., & Moritz, P. (2003). An alternative paradigm for clinical nursing research: An exemplar. *Research in Nursing and Health, 26,* 244–255.

TenHave, T. R., Coyne, J., Salzer, M., & Katz, I. (2003). Research to improve the quality of care for depression: Alternatives to the simple randomized clinical trial. *General Hospital Psychiatry, 25,* 115–123.

Torgerson, D. J., Klaber-Moffett, J., & Russell, I. T. (1996). Patient preferences in randomised trials: Threat or opportunity? *Journal of Health Services Research and Policy, 1*(4), 194–197.

Tunis, S. R., Stryer, D. B., & Clancy, C. M. (2003). Practical clinical trials: Increasing the value of clinical research for decision making in clinical and health policy. *Journal of the American Medical Association, 290,* 1624–1632.

Vandenbroucke, J. P. (2004). When are observational studies as credible as randomised trials? *Lancet, 363,* 1728–1731.

Whittemore, R., & Grey, M. (2002). The systematic development of nursing interventions. *Journal of Nursing Scholarship, 34,* 115–120.

Souraya Sidani, PhD, RN, is Associate Professor, Faculty of Nursing, University of Toronto, Ontario, Canada.

Résumé

L'évaluation de l'efficacité des interventions : une exploration de deux méthodes statistiques

Mary T. Fox, Angela Cooper Brathwaite et Souraya Sidani

La formule de mesures répétées est souvent utilisée pour évaluer l'efficacité des interventions. Selon cette formule, les résultats sont mesurés à plusieurs reprises, soit avant et après la mise en œuvre du plan d'intervention. Les données peuvent être analysées à l'aide de deux méthodes statistiques : l'analyse de la variance fondée sur le mesurage répété (RM-ANOVA) et le modèle linéaire hiérarchique (HLM). Les auteures offrent un aperçu des modèles statistiques propres à la RM-ANOVA et au HLM et discutent des forces et des limites de chacun d'eux tout en suggérant que ces deux méthodes sont complémentaires dans une démarche visant à mesurer l'efficacité des interventions.

Mots clés : analyse de la variance

CJNR 2004, Vol. 36 N° 3, 20–30

Evaluating the Effectiveness of Interventions: Exploration of Two Statistical Methods

Mary T. Fox, Angela Cooper Brathwaite, and Souraya Sidani

Repeated measures designs are often used to evaluate the effectiveness of interventions. In these designs, the outcomes are measured on several occasions before and after implementation of the intervention. Two statistical methods, the repeated measures analysis of variance (RM-ANOVA) and hierarchical linear models (HLM), can be used to analyze the data. The authors provide an overview of the statistical models underlying RM-ANOVA and HLM and discuss the strengths and limitations of each. They propose that the 2 methods are complementary in determining the effectiveness of interventions.

Keywords: research design, treatment outcome, outcome assessment (health care), analysis of variance, regression analysis, linear models, longitudinal studies, nursing research

Background

The provision of empirical evidence regarding the effectiveness of interventions is critical for decision-making in clinical, administrative, and educational nursing practice. Appropriate implementation requires knowledge about the extent to which an intervention is beneficial. Such knowledge is usually derived from the results of studies that evaluated the effects of an intervention on the intended outcomes. Intervention evaluation studies tend to use repeated measures designs. A repeated measures design is one in which the same outcome variables are measured for each participant at several points in time (Daniel, 1999; Keselman, Algina, & Kowalchuk, 2001). The study involves one or more groups. The points in time are selected to represent the participants' status before and after implementation of the intervention. The statistical analysis is aimed at determining the extent to which the participants' standing on the outcomes changed as a result of the intervention.

Two statistical methods, repeated measures analysis of variance (RM-ANOVA) and hierarchical linear models (HLM), can be used to analyze the data obtained from intervention studies with repeated measures designs. These methods provide different yet complementary information about the effectiveness of the intervention. The findings of RM-ANOVA

indicate whether the mean scores on the outcome differed across occasions of measurement, within the same group or between groups (e.g., experimental and control groups). These findings, presented at the group level, are informative. However, they may not be sufficient to guide decision-making in practice; nurses need to know the percentage of participants who demonstrated the anticipated pattern of change in the intended outcome (Jacobson & Traux, 1991; LeFort, 1993; Sidani & Braden, 1998). HLM is a complementary data analytic method. It generates findings showing the pattern of change in the outcome for individual participants and the percentage of participants who demonstrated the anticipated improvement.

In this paper, we briefly review the statistical models underlying RM-ANOVA and HLM and discuss the strengths and limitations of each method in intervention evaluation research. We then present an empirical example to illustrate the complementary nature of the results obtained using the two methods; the results provide a comprehensive depiction of the effectiveness of the intervention using both methods.

Repeated Measures Analysis of Variance (RM-ANOVA)

RM-ANOVA is used to assess the effectiveness of an intervention in studies with repeated measures of the outcomes. The outcomes are measured at pre-test, post-test, and follow-up intervals in either the same group of participants who received the intervention, or two or more groups of participants who did or did not receive the intervention — that is, experimental or control groups, respectively. In experimental research, it is assumed that the intervention effects show up as a change of a constant value in the participants' post-test outcome scores (Lipsey, 1990; Raudenbush & Bryk, 2002). Therefore, the data analysis focuses on determining whether the mean scores on the outcome differ between groups and/or across occasions of measurement.

RM-ANOVA is a "multivariate inferential statistical procedure that is used to compare performance on a single dependent variable at three or more points in time" (Norwood, 2000, p. 364). The statistical model underlying RM-ANOVA is an extension of the general linear model. It tests for group, time, and group-by-time interaction effects, depending on the number of groups included in the study. The appropriate F ratio is computed to determine whether each effect is statistically significant. A statistically significant group effect indicates that the mean scores on the outcome differ between the experimental and control groups; however, it gives no clear indication of the measurement occasion(s) at which the between-group difference occurred. Post hoc analyses using independent sample t tests are conducted to compare the means of the two groups at

each point in time. A statistically significantly time effect implies that the mean scores on the outcome for the total sample differed across occasions of measurement, but there is no clear indication of when exactly the difference was noticed. Post hoc analyses using paired *t* tests are conducted to compare the means observed at the different occasions of measurement. In the latter case, the mean comparisons between occasions are done within each group as well as for the total sample (Green, Salkind, & Akey, 1999). A statistically significant group-by-time interaction effect indicates that the mean scores on the outcome differed between groups and across occasions of measurement. Post hoc comparisons are conducted to identify the point(s) in time at which the groups differed (Green et al.; Littell, Henry, & Ammerman, 1998).

The *F* ratio used in the RM-ANOVA and the *t* test used in the post hoc comparisons examine differences in the outcome at the group level. The results are presented for the "average" participant (Barlow, 1996) in terms of an "average" amount of change in the outcome between adjacent points in time (Hartman, Stage, & Webster-Stratton, 2003; Wu, Clopper, & Wooldridge, 1999). The emphasis on the group-level analysis and the mean change in the outcome may not be sufficiently informative to guide practice (Jacobson & Truax, 1991) and may be potentially misleading (Francis, Fletcher, Stuebing, Davidson, & Thompson, 1991). The *F* ratio and the post hoc *t* tests, and their associated results, do not provide any information either on the variability in the outcome achieved by the participants or on the pattern of change in the outcome across occasions of measurement. The participants may have differed in their responses to the intervention; some may have fully benefited, others not benefited at all, and others benefited to varying degrees (Brown, 2002; Sidani & Epstein, 2003).

In the statistical model underlying the RM-ANOVA *F* ratio and the post hoc *t* tests, the individual variability in the outcome is reflected in the within-group variance, which is considered the error term (Francis et al., 1991). Increased within-group variance reduces the statistical power to detect significant differences in the outcome between groups and/or across occasions of measurement, which in turn increases the likelihood of committing a type II error — that is, erroneously concluding that the intervention was not effective (Lipsey, 1990). Therefore, it is critical to know how many participants demonstrated the anticipated improvement in order to determine the benefits derived from the intervention (Jacobson & Truax, 1991; LeFort, 1993).

The results of RM-ANOVA and post hoc comparisons indicate that the mean scores differed over time, suggesting a certain amount of change in the group's level on the outcome. These findings do not provide information on the pattern of change — that is, the direction

and magnitude of the change in the outcome (Bryk & Raudenbush, 1987). Trend analysis can be performed to determine the pattern of change in the outcome; however, this analysis focuses on the average change observed for the group, not individual variability in the pattern of change (Francis et al., 1991; Wu et al., 1999). Evidence regarding the number of participants who changed and their patterns of change not only is clinically relevant but also may minimize the potential for drawing inaccurate conclusions regarding the effectiveness of the intervention.

Hierarchical Linear Models (HLM)

HLM, also known as growth curve or individual regression analysis, is a statistical technique that can be used to analyze individual differences in the pattern of change in the outcome (Floreck & De Champlain, 2001; Willett, Singer, & Martin, 1998). Its results complement those obtained from RM-ANOVA; the results derived from HLM describe the direction and magnitude of change in the outcome for each participant and indicate whether the pattern of change differed between the study groups (Raudenbush & Bryk, 2002; Warschausky, Kay, & Kewman, 2001). The assumption underlying HLM is that change is a continuous process, which is best described by a trajectory rather than a series of discrete alterations observed at fixed points in time (Francis et al., 1991; Wu et al., 1999). The focus is on modelling, describing, and explaining the variability in the trajectory or pattern of change.

To clarify the statistical model of HLM, we present it in two steps, the first describing the analysis conducted at the individual level, the second explaining the analysis conducted at the group level. The second step is performed if two or more groups are included in the study.

In the first step, the analysis estimates the pattern of change in the outcome measured before and after implementation of the intervention for each participant (Lipsey & Cordray, 2000). In this analysis, the data should be entered for each participant by occasion of measurement, as detailed by Raudenbush and Bryk (2002) and Sidani and Lynn (1993). A time variable is created to reflect the different points of measurement, which are assigned consecutive numeric values. The outcome variable is created to reflect the score obtained at each point of measurement. The outcome variable is regressed on the time variable, for each individual. The regression equation is expressed as $Y_i = B_{oi} + B_{1i}$ (Time) + error. Y_i represents the outcome variable for each participant. B_{oi} is the intercept; it describes each individual participant's value on the outcome at pre-test. B_{1i} is the slope; it describes the pattern — that is, the direction and magnitude — of change in the outcome for each participant. Time

represents the variable reflecting the different points of measurement. Error is the random error of prediction.

Each participant's slope is examined for its direction. A negative slope indicates that the participant's level on the outcome decreased, whereas a positive slope implies that it increased over time. Also, the standardized slope is evaluated for its magnitude. A slope of zero indicates that no change occurred in the outcome over time. A slope of less than .30 indicates that the participant's level on the outcome showed a small, gradual change over time. A slope between .30 and .60 indicates that the participant's level on the outcome showed a moderate but steady change. A slope greater than .60 indicates a large, rapid change in the outcome. Visual examination of the individual regression lines is highly recommended, to determine whether some participants exhibited non-linear, such as inverted ∩ or S-shaped, patterns of change. For these participants, the regression model should incorporate the appropriate terms (e.g., polynomials) that account for non-linearity (Kleinbaum, Kupper, & Muller, 1988; Warschausky et al., 2001).

In the second step of HLM, the analysis is intended to determine the extent to which the pattern of change in the outcome varied between the groups of participants. This is done with regression-type analyses. The individual participants' slopes (B_{1i}) that were obtained in the first step are regressed onto the variable(s) hypothesized to influence the pattern of change in the outcome. These may include: (1) the participant's level on the outcome measured at baseline or pre-test, which is represented by the intercept (B_o) obtained in the first step — the baseline outcome level is usually included in the regression analysis when significant inter-individual differences are observed at baseline; (2) the group to which the participant was assigned or the dose of the intervention to which the participant was exposed, as suggested by Sidani (1998). The latter variable is used when a dose-response analysis is conducted. The regression weights or beta parameters associated with each variable are examined for magnitude, direction, and statistical significance, as they indicate the extent to which these variables influenced the pattern of change in the outcome (Raudenbush & Bryk, 2002; Tate & HoBanson, 1993; Warschausky et al., 2001).

The results of HLM point to the direction and magnitude of change in the outcome exhibited by the participants from pre-test to post-test or follow-up. Variability in the pattern of change is acknowledged and considered of importance and interest. The number and percentage of participants who showed no change or different patterns of change are reported and may indicate the extent to which the intervention was beneficial. Intervention benefits can be inferred if a large percentage of par-

ticipants who received it showed the anticipated direction and magnitude of change in the outcome.

Illustrative Example

The data used in the following illustrative example were obtained from a study that evaluated the effectiveness of a staff-development educational program aimed at enhancing nurses' cultural knowledge and competence (Cooper Brathwaite, 2004). The program consisted of five sessions during which the theory and principles underlying cultural competence were discussed and the skills required for providing culturally competent care were reinforced. The program was offered to 76 public health nurses.

A one-group repeated measures design was used to evaluate the effects of the program. This design, which ascertains pattern of change over time in the same group of individuals and change within each individual under control and post-intervention conditions (Burns & Grove, 2001), was selected for several reasons. By not having separate treatment and control groups, it avoided dissemination of the intervention beyond the treatment group and compensatory rivalry among nurses working at the same site (Cooper Brathwaite, 2004). Also, the design minimized the potential for selection bias since the intervention was not given to nurses working at one site and withheld from those working at another (Cooper Brathwaite).

The outcome data were collected at four occasions separated by equal time intervals of 3 months. The first two occasions (time 1: first pretest; time 2: second pretest) represented the control condition. Time 3 data were obtained following implementation of the program. Time 4 data reflected the 3-month follow-up, which was used to determine the sustainability of the program's effects. The outcome variables were measured with self-report instruments that demonstrated acceptable reliability and validity. Cultural knowledge was measured using the Cultural Knowledge Scale (CKS) adapted from the cultural knowledge and cultural efficacy subscales developed by Campinha-Bacote (1999) and Bernal and Froman (1993), respectively. Cultural competence was measured using Campinha-Bacote's Inventory for Assessing the Process of Cultural Competence (IAPCC).

The RM-ANOVA was used to determine whether the group's mean scores on each outcome differed across occasions of measurement. A significant time effect [$F(3, 69) = 142.02$, $p < .01$)] was found for CKS. The post hoc comparisons showed that: (a) the mean score at time 1 ($3.76 \pm .26$) did not differ significantly from the mean score at time 2 ($3.77 \pm .27$); (b) the mean score at time 2 was significantly different from the mean score at time 3 ($4.57 \pm .34$); and, (c) the mean score at time 3

did not differ from that observed at time 4 (4.58 ± .39). Similarly, a significant time effect [$F(3, 69) = 118.87, p < .01$] was found for IAPCC. The results of the post hoc comparisons indicated that the mean scores differed between time 1 (2.87 ± .23) and time 2 (2.82 ± .18), time 2 and time 3 (3.38 ± .34), and time 3 and time 4 (3.51 ± .37).

The data were also analyzed using HLM in order to describe the pattern of change in each outcome within the group. The slope was estimated for each participant. For CKS, the estimated slope values ranged from .02 to .59, with a mean of .32 (± .14). Specifically, four participants (5.6%) had a slope close to zero (0 to .10), reflecting no gain in their cultural knowledge over time. For 47.2% of the participants, the slope varied between .11 and .30, reflecting a small, gradual gain in knowledge. For the remaining 47.2% of the participants, the slope ranged from .31 to .60, implying a moderate but steady gain in cultural knowledge. For IAPCC, the estimated slope value ranged between .00 and .53, with a mean of .23 (± .12). About 18.1% of the participants had a slope value of .00 to .10; 55.5% had a slope value of .11 to .30; and 26.4% had a slope value of .31 to .60.

The findings of the RM-ANOVA indicated that (a) the intervention was effective in increasing the nurses' level of cultural knowledge and competence (evidenced by the significant difference between the mean scores at times 2 and 3), and (b) the level of knowledge attained at posttest was maintained at 3-month follow-up, while the level of competence continued to increase at follow-up, as hypothesized. Although these results support the effectiveness of the intervention, they do not inform us of how many participants showed the anticipated improvement in the outcomes. The findings of the HLM clarified that approximately half of the participants increased their cultural knowledge and about a quarter demonstrated improvement in their level of cultural competence.

Conclusion

The clinical usefulness and applicability of results derived from intervention studies rests on the researcher's ability to provide evidence on the nature of change that results from an intervention. The analysis of longitudinal data derived from intervention studies has traditionally been conducted using RM-ANOVA. As such, the description of change has been chiefly limited to whether or not there was a change for the average participant from the target population. Although this is valuable information, it is insufficient as a basis for clinical, educational, and administrative decisions in nursing. Knowing the extent to which participants changed and their specific patterns of change is requisite for evidence-based decision-making. This level of information is not provided by RM-ANOVA.

Therefore, nurse researchers need to expand their repertoire of data-analytic approaches. HLM is an approach that researchers can use to complement RM-ANOVA in the analysis of longitudinal data derived from intervention studies.

Used together, RM-ANOVA and HLM equip researchers to provide more complete information on the effectiveness of interventions. RM-ANOVA can provide information on differences in change between groups. HLM can complement this level of information by describing the pattern of change, how various subgroups and/or individuals of the population changed, the proportion of individuals who changed, and the ways in which they changed in response to the intervention. For example, the information presented in the illustrative example using RM-ANOVA indicated that the educational program was effective, on average, in increasing public health nurses' cultural knowledge and competency. Although this information can be helpful to the nurse administrator in deciding whether to implement the educational program, it cannot be used to infer the efficacy or utility of the program. However, using the additional information obtained from the HLM analysis, the administrator can anticipate that the educational intervention will benefit approximately half of the public health nurses in terms of increasing their cultural knowledge and approximately one quarter in terms of increasing their cultural competence. Moreover, the nurse administrator can use the results from the HLM analysis to anticipate that 47% of the nurses who attend the educational program will have a moderate steady increase in cultural knowledge and 26% will have a moderate steady increase in cultural competence over a 3-month period. Furthermore, this additional information can be extended to estimate the cost effectiveness of the program.

References

Barlow, D. H. (1996). Health care policy, psychotherapy research, and the future of psychotherapy. *American Psychologist, 51,* 1050–1058.

Bernal, H., & Froman, R. (1993). Influences on the cultural self-efficacy of community health nurses. *Journal of Transcultrual Nursing, 4*(2), 24–31.

Brown, A. J. (2002). Nursing intervention studies: A descriptive analysis of issues important to clinicians. *Research in Nursing and Health, 25,* 317–327.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*(1), 147–158.

Burns, N., & Grove, S. K. (2001). *The practice of nursing research*. Toronto: W. B. Saunders.

Campinha-Bacote, J. (1999). A model and instrument for addressing cultural competence in health care. *Journal of Nursing Education, 38*(5), 203–207.

Cooper Brathwaite, A. (2004). *Evaluation of a cultural competence educational intervention*. Unpublished doctoral dissertation, University of Toronto.

Daniel, W. (1999). *Biostatistics: A foundation for analysis in the health sciences* (7th ed.). Toronto: John Wiley.

Floreck, L. M., & De Champlain, A. E. (2001). Standardized patients – will the questions never end? *Academic Medicine, 76*(10), S93–S95.

Francis, D. J., Fletcher, J. M., Shuebing, K. K., Davidson, K. C., & Thompson, N. M. (1991). Analysis of change: Modeling individual growth. *Journal of Consulting and Clinical Psychology, 59*(1), 27–37.

Green, S., Salkind, N., & Akey, T. (1999). *Using SPSS for Windows: Analyzing and understanding data* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Hartman, R. R., Stage, S. A., & Webster-Stratton, C. (2003). A growth curve analysis of parent training outcomes: Examining the influence of child risk factors (inattention, impulsivity, and hyperactivity problems), parental and family risk factors. *Journal of Child Psychology and Psychiatry, 44*(3), 388–398.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12–19.

Keselman, H. J., Algina, J., & Kowalchuk, R. K. (2001). The analysis of repeated measure designs: A review. *British Journal of Mathematical and Statistical Psychology, 54,* 1–20.

Kleinbaum, D. G., Kupper, L. L., & Muller, K. E. (1988). *Applied regression analysis and other multivariable models* (2nd ed.). Boston: PWS-KENT.

LeFort, S. M. (1993). The statistical versus clinical significance debate. *Image: Journal of Nursing Scholarship, 25*(1), 57–62.

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Thousand Oaks, CA: Sage.

Lipsey, M. W., & Cordray, D. S. (2000). Evaluation methods for social intervention. *Annual Review of Psychology, 51*, 345–375.

Littell, R. C., Henry, P. R., & Ammerman, C. B. (1998). Statistical analysis of repeated measures data using SAS procedures. *Journal of Animal Science, 76*(4), 1216–1231.

Norwood, S. L. (2000). *Research strategies for advanced practice nurses.* Upper Saddle River, NJ: Prentice Hall.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Sidani, S. (1998). Measuring the intervention in effectiveness research. *Western Journal of Nursing Research, 20*(5), 621–635.

Sidani, S., & Braden, C. J. (1998). *Evaluating nursing interventions: A theory driven approach*. London: Sage.

Sidani, S., & Epstein, D. R. (2003). Enhancing the evaluation of nursing care effectiveness. *Canadian Journal of Nursing Research, 35*(3), 26–38.

Sidani, S., & Lynn, M. R. (1993). Examining amount and pattern of change: Comparing repeated measures ANOVA and individual regression analysis. *Nursing Research, 42*(50), 283–286.

Tate, T. L., & HoBanson, J. E. (1993). Analyzing individual status and change with hierarchical linear models: Illustration with depression on college students. *Journal of Personality, 61*(2), 181–206.

Warschausky, S., Kay, J. B., & Kewman, D. G. (2001). Hierarchical linear modeling of FIM instrument growth curve characteristics after spinal cord injury. *Archives of Physical Medicine Rehabilitation, 82,* 329–334.

Willett, J. B., Singer, J. D., & Martin, N. C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations. *Development and Psychopathology, 10,* 395–406.

Wu, Y. W. B., Clopper, R. R., & Wooldridge, P. J. (1999). A comparison of traditional approaches to hierarchical linear modeling when analyzing longitudinal data. *Research in Nursing and Health, 22,* 421–432.

Authors' Note

Comments or inquiries may be directed to Mary Fox, Director, Collaborative Research Program, Rehabilitation and Long-Term Care, Nursing Services Administration, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, Toronto, Ontario M6A 2E1 Canada. Telephone: 416-785-2500, Ext. 2714. Fax: 416-785-2501. E-mail: mfox@baycrest.org

---

Mary T. Fox, RN, MSc, is a PhD candidate in the Faculty of Nursing, University of Toronto, Ontario, Canada; Director, Collaborative Research Program, Rehabilitation and Long-Term Care, Baycrest Centre for Geriatric Care, Toronto; and a fellow of the Canadian Institutes of Health Research. Angela Cooper Brathwaite, RN, MN, CHE, PhD, is Manager, Public Health Nursing and Nutrition, Durham Region Health Department, Whitby, Ontario; and Lecturer, Faculty of Nursing, University of Toronto. Souraya Sidani, RN, PhD, is Associate Professor, Faculty of Nursing, University of Toronto.

Résumé

Intersectionalité d'influences : la recherche sur la santé des femmes immigrantes et réfugiées

Sepali Guruge et Nazilla Khanlou

On s'entend aujourd'hui de plus en plus pour reconnaître la complexité qui caractérise de multiples axes ou dimensions de l'identité sociale et la façon dont ceux-ci se recoupent pour influer sur la santé des femmes immigrantes et réfugiées. Le concept d'intersectionalité d'influences est particulièrement approprié pour aborder la diversité dans le cadre de la recherche en sciences infirmières. L'objectif du présent article est d'élaborer une théorie sur ce concept et de le rendre opérationnel dans la recherche visant la promotion de la santé mentale chez les femmes immigrantes et réfugiées. Sur le plan conceptuel, les auteures proposent une méthode d'enquête qui est influencée par le savoir critique et qui puise à même la pensée post-coloniale et féministe. Sur le plan opérationnel, elles appliquent un cadre écosystémique afin de déterminer quels sont les problèmes de santé individuels dans les sphères familiale, communautaire et sociale. Elles présentent la recherche-action participative comme un moyen de mettre ces concepts en œuvre dans le processus de recherche. Leur objectif consiste à introduire un nouveau mode d'enquête qui s'avérera avantageux pour les femmes immigrantes et réfugiées tout en faisant progresser l'approche des sciences infirmières dans la recherche communautaire.

Mots clés : post-colonial, féministe, cadre écosystémique, recherche-action participative, santé des femmes immigrantes et réfugiées, intersectionalité d'influences.

CJNR 2004, Vol. 36 N° 3, 32–47

Intersectionalities of Influence: Researching the Health of Immigrant and Refugee Women

Sepali Guruge and Nazilla Khanlou

There is a growing recognition of the complexity surrounding multiple axes or dimensions of social identity and how they intersect to influence the health of immigrant and refugee women. The concept of intersectionalities of influence is particularly relevant in addressing diversity in nursing research. The purpose of this paper is to theorize and operationalize the concept in mental health promotion research with immigrant and refugee women. At the conceptual level, the authors propose an approach to inquiry that is informed by critical scholarship and draws from postcolonial and feminist perspectives. At the operational level, they apply an ecosystemic framework to help locate individual health issues within the familial, community, and social realms. They introduce Participatory Action Research as a way of putting these concepts into action within the research process. Their aim is to introduce a new way of inquiry that can benefit immigrant and refugee women while furthering the nursing agenda for community-based research.

Keywords: postcolonial, feminist, ecosystemic framework, Participatory Action Research, immigrant and refugee women's health, intersectionalities of influence

As communities in Canada grow in diversity, nurse researchers in health promotion are faced with the responsibility of addressing the needs of all communities. We posit that the exploration of immigrant and refugee women's health requires new ways of inquiry, ones that move beyond the traditional ways of conceptualizing and operationalizing research. New ways of inquiry should create space for the exploration of how various dimensions of social identity, such as race, gender, and class, as well as education, citizenship, and geographical locations, intersect to influence the health of immigrant and refugee women. These dimensions are transformed into hierarchies and get built into institutional structures, legitimizing inequalities among different groups of women (Dhruvarajan & Vickers, 2002). Accordingly, new ways of inquiry into the health issues and concerns of immigrant and refugee women must locate individual health and illness experiences within the complex socio-economic, historical, political, and institutional structures and dynamics in the pre- and post-migration context.

The purpose of this paper is to introduce a new approach to examining differences arising from the intersectionality of various dimensions

of social identity that can place women in various locations in the hierarchical social space, and the impact of this on the health of immigrant and refugee women. Our intention is not to overemphasize differences or to minimize the shared values, beliefs, experiences, and expectations of a particular ethnocultural group or the common experiences of immigration or of being a woman.

An interesting emergent concept in the literature is "intersectionalities of influence." In the health context, this alludes to the intersection of multiple sources of influence on our physical, mental, and spiritual health and well-being. The concept recognizes that influences come together in distinct ways, leading to different health outcomes for individuals or groups (Khanlou, 2003). Thus it contextualizes individual health experiences and pays heed to the combination of forces arising from *micro*-, *meso*-, and *macro*-levels of society. We posit that the idea can be of particular relevance in mental health promotion research with immigrant and refugee women.

In this paper, we attempt to theorize and operationalize the concept of intersectionalities of influence. To this end, we propose a theoretical stance that draws from postcolonial and feminist theoretical perspectives. Such theoretical perspectives provide a lens through which to view issues and to understand and explain social reality. However, we require a framework, a conceptual link, in order to capture the complexity of the proposed theoretical domains. A framework has been described as an "abstract, logical structure of meaning that guides the development of the study and enables the researcher to link the findings to nursing's body of knowledge" (Burns & Grove, 2001, p. 44). We believe the ecosystemic framework serves this purpose. We then introduce Participatory Action Research (PAR) as a way of putting the above concepts into action within the research process. Our purpose here is not to create a unified field but to explore a new and different way of thinking. We draw on the theoretical literature and discuss various examples from our own research and that of others to show that such a method of inquiry can benefit immigrant and refugee women while furthering the nursing research agenda.

Postcolonial Feminist Theoretical Perspectives in Nursing Research

There has been a call for and move towards critical inquiry in nursing. Informed by critical traditions such as critical social theory, cultural studies, lesbian/gay studies, and feminist studies, "contemporary nursing discourses draw our attention to the convergence of several factors within the specific domain of nursing science and the larger arenas of

health care and social inquiry" (Reimer-Kirkham & Anderson, 2003, p. 2). Critical inquiry allows us to move beyond simple representation of the world, to uncover hidden assumptions by challenging the whole process of framing and doing research and explore issues of importance to all women by taking into account both inter- and intra-group diversity. It involves the questioning of assumptions and taking a more critical and reflexive approach to the examination of our positionality in the context of our work. Critical inquiry helps us to reflect on how production, reproduction, and presentation of knowledge within a particular ideological foundation can not only perpetuate existing power relations, inequity, and vulnerability, but also result in the further marginalization of racialized women (Anderson, 2002; Tuhiwai Smith, 2001).

There is also emerging interest in the integration of postcolonial perspectives in nursing (Anderson, 2002; Anderson et al., 2003; Reimer-Kirkham & Anderson, 2003; C. Varcoe, personal communication, March 2004). Postcolonial scholarship is the theoretical and empirical examination of issues stemming from colonial relations and their aftermath (Cashmore, cited in Reimer-Kirkham & Anderson). "Its concern extends to the experiences of people descended from the inhabitants of those territories and their experiences within 'first-world' colonial powers" (Reimer-Kirkham & Anderson, p. 3). Postcolonial theoretical perspectives help us to explore the effects of intersecting forces of influence in the everyday lives of women in the post-migration context, and how they understand and interpret their health and settlement concerns from their location in the hierarchical social space. Further, a postcolonial theoretical stance allows us to understand how the views of immigrant and refugee women are shaped by their experiences in their own countries, their lives during displacement, and their experiences within exiled communities in Canada. In addition, such a stance allows us to see that these views are not discrete but exist in dynamic interaction with one another.

In particular, postcolonial theoretical perspectives provide us with an "analytic apparatus to examine how the 'non-Western Other' has been constructed through contrasting images with the West" (Anderson, 2002, p. 12). This theoretical lens also reveals "the everyday experiences of marginalization, as structured by the micropolitics of power and the macrodynamics of structural and historical nature" (Reimer-Kirkham & Anderson, 2003, p. 2). Only if we seriously consider such a perspective and, as Tuhiwai Smith (2001) argues in the context of research and indigenous peoples, research back and disrupt "the rules of the research game toward practices that are more respectful, ethical, sympathetic, and useful vs. racist practices, attitudes, and ethnocentric assumptions and exploitative research" (Preface), can we produce knowledge that is truly meaningful and relevant for immigrant and refugee women in Canada.

A postcolonial theoretical approach on its own does not necessarily include a gendered perspective; for that we turn to feminist scholarship. (See Anderson [2002] and Reimer-Kirkham and Anderson [2003] for a detailed discussion of postcolonial feminism and postcolonialism, respectively.)

Feminist scholarship has been widely acknowledged as having legitimacy for breaking new ground and creating a space for contesting androcentric and positivistic inquiry into women's concerns and issues. Common to most feminist work is the notion that feminist research is for women, by women, and that its purpose is to learn about women and their experiences, making visible what is important to women while giving them a voice (Harding, 1987, 1989; Kirsch, 1999; Seibold, 2000). However, various nursing scholars have begun to criticize "Western feminism" for the narrowness it imposes on an applied discipline such as nursing (Peter, 2000; Seibold; Thorne & Varcoe, 1998). According to Dhruvarajan and Vickers (2002), until recently many feminist theories and practices "did not deal well with forces such as racism, nationalism, class conflict and homophobia, and ablism. Nor did many mainstream feminists understand why some women world wide are sceptical of feminism and are likely to be mobilized by movements dedicated to nationalism, socialism, antiracism, or gay rights, or even by antifeminist movements, and the fundamentalist movements now associated with many religions" (p. 6). Seibold also argues that the uncritical assumption of "a necessary bond between being a woman and occupying certain social roles does not help to uncover the ways in which women negotiate the world and the wisdom inherent in such a negotiation" (p. 152). We need to promote critical feminist approaches in order to challenge and resist the status quo, demanding representation of all women and their issues in nursing research. Critical feminist approaches should not only be sensitive to "the dominant culture's devaluation of caring and nurturing practices, like caring for the sick and dying, [and] mothering" (Peter, p. 102), but also examine why the values of non-dominant cultural groups (such as collectivism versus individualism, or various hybrids of the two) are often held in less regard. Unless explicit attention is given to the underlying values of our research, the resulting knowledge may be used to change non-dominant cultural groups so that they accept dominant values, beliefs, and practices. Embedded in such knowledge is the assumption of Western superiority.

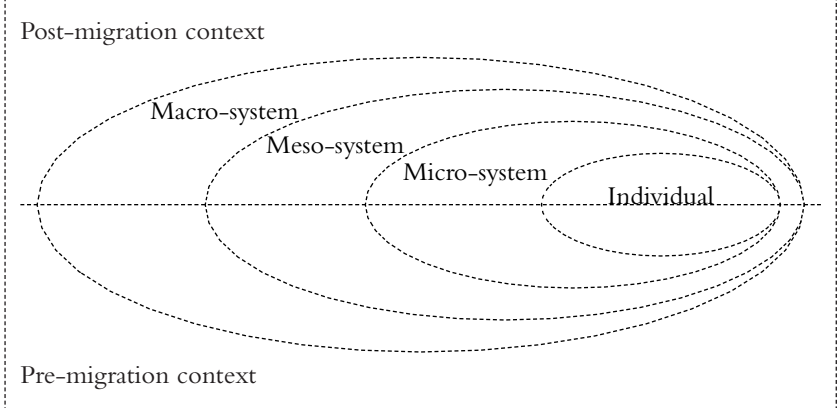
A postcolonial feminist perspective allows us to examine the complex issues at the intersection of gender, race, class, and culture, and move beyond single causal factors of illness or *micro*-level analysis. It directs us to see how the complex historical, political, cultural, and socio-economic context in which human experience is embedded affects and shapes the

health and illness experiences of immigrant and refugee women and their access to equitable and quality care. A postcolonial feminist theoretical perspective "recognizes the need for knowledge construction from the perspective of the marginalized female subject whose voice has been muted in the knowledge production process" (Reimer-Kirkham & Anderson, 2003, p. 10). It permits us to see how some of the systemic practices and structural barriers in the host country reduce the space for resistance and hope on the part of immigrant and refugee women, and that migration to a country of the North does not necessarily improve a woman's status both within and outside the home.

Ecosystemic Framework

Although it is applied in such disciplines as social work and psychology, the ecosystemic framework has not received much attention in nursing. We posit that the framework helps to operationalize the concept of intersectionalities of influence in researching the health issues and concerns of immigrant and refugee women. It also helps to put into research practice the complex theoretical domain of postcolonial feminism. The ecosystemic framework (often used interchangeably with the ecological framework) considers individual situations as arising from the transaction between the *individual*, family (*micro*-system/level), community (*meso*-system/level), and larger social and cultural environment (*macro*-system/level); the transactions between the four systems are seen as continuous and reciprocal (Germain & Bloom, 1999; Waller, 2001). A visual representation of the framework, a modification of Heise's (1998) version as noted in Guruge (2004), is provided in Figure 1.

Figure 1 ***Ecosystemic Framework***



Here, we introduce a few examples of the factors that fall under each of the four levels/systems. At the *individual* level, immigrant and refugee women may define themselves in ways that acknowledge, to varying degrees, their race, ethnicity, class, gender, age, language, education, and citizenship status. Other individual factors include childhood experiences, biology, and genetics (Germain & Bloom, 1999; Heise, 1998). The intersections of these dimensions create differences both within and between groups of women. Often in health care, self is seen to exist as an isolated entity and illness as a result of internal deficits or conflicts, and therefore treatments are directed towards the individual. While acknowledging that illness may occur due to individual biophysiological pathology, we highlight the fact that the individual receives and engages in care and treatment in a context of interactions with others. At the *micro*-level is the family (Germain & Bloom; Heise). The health and well-being of immigrant and refugee women is often closely linked to that of their family members and shaped by their relationships with them, whether at the local, national, and/or transnational level. However, the family, both immediate and extended, can be a source of both strength and stress for women and thus can have a powerful influence on the individual. The informal and formal social networks that form the community make up the *meso*-level. Forces within one's social network, such as school, neighbourhood, and workplace, influence the individual both positively and negatively. For example, Anderson, Blue, Holbrook, and Ng (1993) found that women in the lower echelons of the workforce have very little opportunity to manage a chronic illness such as diabetes. *Macro*-level factors include the "health, educational, economic and social policies that maintain high levels of economic or social inequality between groups in society" (Krug, Dahlberg, Mercy, Zwi, & Lozano, 2002, p. 13). There is a growing body of evidence that considers these factors under the category of social determinants of health (Wilkinson & Marmot, 1998). For example, various immigration policies as well as ethnocentric health-care facilities can limit the choices refugee women have in dealing with intimate partner abuse in a manner that is least damaging to themselves and their families. Although there is space for interpretation and debate as to exactly where a particular factor might fit within the framework, more important than the location of any single factor is the dynamic interplay between factors operating at multiple levels.

The ecosystemic framework provides a basis for analysis of the complex issues at the intersection of race, gender, and class identities (both imposed and assumed) and the interaction of these and other identities with *micro*-, *meso*-, and *macro*-level factors and issues. Only by examining structural and systemic inequalities, and their impact on the

health and well-being of immigrant and refugee women, can we hope to achieve equitable health care for all women.

From Theory to Action

The demand for objectivity in the research process in nursing and other health-care professions disregards the role that researchers play in the construction of knowledge. Regardless of the paradigm of research and the design tradition employed, it is the researchers who decide what to research, how to frame the research questions, who to interview, how to access study participants, what questions to ask, how to probe the answers, how to analyze data, how the findings are used, and so forth (Clough, 1992, 1994). According to Duster (2000), one of the key questions in knowledge production is "whose questions get raised for investigation" (p. xii). We also need to reflect on the question Whose interests are we serving by doing research that will not be directly beneficial or applicable to the participants and participant communities? A reflexive, participatory, collaborative, and collective approach to research allows community members themselves to articulate what is important to the community and facilitates the generation of knowledge that will enhance the capacity of both individuals and communities in the post-migration and settlement context.

In the next section we introduce PAR as a way of putting the ideas discussed earlier in the paper into action within the research process.

Participatory Action Research

We find PAR particularly relevant for our field of research in nursing, as there are many overlapping conceptual points of convergence between mental health promotion and PAR. For example, both recognize the capacities of people, seek to effect improvements in their lives, and emphasize action (Khanlou et al., 2002). PAR provides a research orientation that is "potentially compatible with feminist research in health promotion" (Denton, Hajdukowski-Ahmed, O'Connor, & Zeytinoglu, 1999, p. 17). Unlike the traditional sequential ordered research process, PAR is cyclical in approach, and the participants themselves are in control of the research process. PAR takes a self-reflexive approach to research: relations of authority and power between the researcher and the participants are taken into account at every step of the study, with the aim of building on trust and respect on the path of collective discovery. Both the researcher and the participants are experts and learners. While "traditional research was detached and 'objective' [PAR] would be engaged, passionate and consciously partial" (Vickers, 2002, p. 68); it "would guard against exploitation of women's time and labour; and it

would work towards structural change" (Denton et al.), an approach that is very much in line with postcolonial feminist perspectives.

PAR provides a space for collective voice, action, and dialogue. Dialogue makes it possible "to reframe dominant discourses and to create a perspective on knowledge development that reflects multiple social locations" (Anderson, 2002, p. 18). The knowledge produced is validated, shared, and used. Such an approach can increase our understanding of not only what is important to various groups and communities, but also the different intersections of influence arising from the structural and systemic barriers that face immigrant and refugee women. It can, thus, reveal to us women's different reactions to the same issues, depending on the location of each woman in the social space. Furthermore, PAR provides a window on "where the power lies and where there is systemic disadvantage, failure to advocate or merit not being recognized or acknowledged" (Hagey, 1997, p. 2). Finally, knowledge gained will be used to increase women's "options for concrete actions, their autonomy in using these options and their capacity to deliberate about choices for action" (Conchelos, cited in McDonald, 2000).

Inherent in a PAR approach is the centrality of voice through group narratives. Yet PAR goes further; it demands agency of all who are involved in the process. We believe these two elements of PAR — voice and agency — make it particularly congruent with a postcolonial feminist theoretical position on research. The centrality of voice is congruent with the long tradition of feminist discourse, dating to Gilligan's (1982) seminal work on women's psychological development. The agency component recognizes the ability of individuals to bring about positive change in their lives, while "a feature of postcolonial scholarship is situating human experience (e.g., everyday reality) in the larger contexts of mediating social, economic, political, and historical forces" (Reimer-Kirkham & Anderson, 2003, p. 11). In addition, arising from action research and participatory research, PAR's distinct geohistorical roots result in a complementary tension, leading to creative health promotion initiatives in transnationalizing societies (Khanlou & Peter, in press).

PAR is an orientation to research rather than a specific research method (Minkler & Wallerstein, 2003). Therefore, depending on the issue under study, it can entail quantitative, qualitative, or combined data-gathering methods. As an approach to conducting research with marginalized populations, PAR brings together the elements of power, investigation, equity, and social justice. Because there is no one single way of conducting participatory research (Israel et al., 2003), there are variations in the degree to which studies are influenced by a PAR approach (Khanlou et al., 2002). We are aware that our discussion of PAR may provide an overly optimistic representation. As with other approaches to research,

PAR has its limitations, which have been addressed elsewhere (Khanlou & Peter, in press).

Case Examples

Drawing from our work, we provide three examples of health research with immigrant youth and women to elucidate our attempt to incorporate the above theoretical concepts in our work. The first two case studies consider the overlap between PAR and mental health promotion (Khanlou et al., 2002; Khanlou & Hajdukowski-Ahmed, 1999). The third examines male violence against women in the post-migration context from a postcolonial feminist perspective, which is operationalized using an ecosystemic framework (Guruge, 2004).

In our studies with immigrant female youth (Khanlou et al., 2002; Khanlou & Hadjukowski-Ahmed, 1999), PAR has provided a model in which the participants' voices are central to the research process. Such a process has revealed to us the contribution of culture to identity development from the perspective of youth in culturally dynamic and immigrant-receiving settings. The framework has also helped to situate the unique experiences of immigrant or second-generation female youth and their families within the multiple systems of influence in the post-migration society of resettlement. Thus our recommendations for mental health promotion among youth recognize the role of systemic challenges as well as individual strengths. The simultaneous health research and health promotion aspects in PAR were revealed in the evaluations by the youth participants at the end of the studies. In one of our studies with newcomer female youth in English-language classes, the youth learned more about others and themselves through their participation in the study's focus groups (Khanlou et al.). For example, one youth reported: "We could know other people's opinion, could find my strength, could find my fear, know more about myself, career planning." Her response and those of others indicate the overlap of self-knowledge, recognition of strengths and challenges, and plans for action. In her proposed doctoral dissertation, Guruge (2004) uses a postcolonial feminist theoretical perspective and an ecosystemic framework to understand male violence against women in the post-migration and resettlement context in Canada. In the framing of the study, domestic violence is seen as arising from the transaction between the individual, family, community, and the larger social and cultural environment within which the couple is embedded. By exploring the pre- and post-migration experiences, one can begin to understand how the intersectionalities of influences at the *micro*-, *meso*-, and *macro*-level of society, both in the couple's country of origin and in the host country, shape what occurs in the home, and how

one man might become abusive towards his wife while another man from the same ethnocultural and religious background would not.

Towards a More Inclusive Research Agenda

Despite increased interest in exploring the various health issues and concerns of immigrant and refugee women in Canada, there are still a number of gaps in the nursing literature. Highlighted here are some gaps — by no means an exhaustive list — that we have come to recognize during our own research and practice. For example, rarely does nursing research explore so-called complementary therapies, which may in fact be the conventional therapies for some women. There is a call for this information in everyday nursing practice (see Guruge, Lee, & Hagey, 2001, for a discussion). Male violence against women is a serious health and social issue; nursing research on violence against women within intimate relationships in the context of post-migration and settlement is just beginning in Canada. The lack of extensive literature on domestic violence in immigrant and refugee communities has far-reaching implications (Guruge, 2004). One implication is the racialization of domestic violence (see Varcoe, 1996, 2001), which leads to the perpetuation and exacerbation of women's vulnerability within the health-care system. Nursing research has paid little attention to women's voices regarding post-migration loss of financial and social stability, the experience of racism and discrimination, the stress of negotiating and navigating through various institutional and structural systems that are designed to serve the dominant groups, and the impact of these experiences on the health and well-being of immigrant and refugee women (see Gastaldo, Khanlou, Massaquoi, & Curling, 2002). The topic of transnational experiences is just beginning to be discussed in the nursing literature. Gastaldo, Gooden, and Massaquoi (2004) argue that "the adjustment of women to a new country cannot be fully understood without an appreciation of the continuing kinship links across national borders." Research into the impact of racism on the health and well-being of immigrant nurses of colour has just begun to appear in the nursing literature (Hagey et al., 2001; Turrittin, Hagey, Guruge, Collins, & Mitchell, 2000). Adaptation, resiliency, and mental health promotion among immigrant youth have begun to be addressed in nursing (Khanlou, 2003; Khanlou et al., 2002). Issues of language translation versus cultural interpretation in providing care have not been researched in nursing even though nurses continue to grapple with them in their daily practice. The idea of cultural safety (Anderson et al., 2003; Hagey, 2000) has not received much attention although there is interest in the topic of patient safety. Nursing in Canada has not paid attention to health issues of

importance to lesbians and bisexual immigrant and refugee women of colour. Peter (2000) posits that the lack of exploration of certain topics in the nursing literature implies that these topics are "of such little importance that they do not require serious reflection and examination" (p. 108).

Nursing needs to also examine the broader issues such as racism, sexism, and classism as well as various institutional and structural elements that continue to support and legitimize inequalities among different groups of women. To this end, we have drawn insights from postcolonial and feminist theoretical perspectives in combination with a PAR orientation. According to the postcolonial perspective, culture is ambiguous, partial, and constructed as a "negotiated process, rather than as static beliefs fixed in time, passed on intact and complete, from one generation to the next" (Anderson, 2002, p. 14). Such an approach provides an opportunity to examine how we might "theorize about culture" in ways that account for shared meanings within groups while leaving an "openness" to shifting identities and realities during diaspora and displacement without contributing to the reinforcement of existing power inequalities (Reimer-Kirkham & Anderson, 2003). Racine (2003) asserts that a postcolonial perspective, when applied to nursing research, "unveils the reductionist Western discourse of essentializing the 'Other' in a unique, crystallized, neutral, rational, and objectivist cultural entity" (p. 96).

Postcolonial feminist PAR offers us a platform from which to explore how "the past, our stories local and global, the present, our communities, cultures, languages and social practices all may be spaces of marginalization, but they have also become spaces of resistance and hope" (Tuhiwai Smith, 2001, p. 4). Such approaches also show us that immigrant and refugee women may have unique ideas as to what is important to them and what needs to be done to transform their lives in health-promoting ways. They also are in congruence with the promotion of "supportive environments and individual resilience, while showing respect for equity, social justice, interconnections and personal dignity" (Centre for Health Promotion, cited in Willinsky & Pape, 1997), and thus promoting the mental health of immigrant and refugee women at the individual and collective levels. Understanding the world of immigrant and refugee women from their vantage point should be on the agenda of nursing research.

Conclusion

In an era of globalization, it is crucial that we make it our priority to address the health needs, issues, and concerns of all women. In order to do so, nursing scholars need to draw from a variety of theoretical approaches. Postcolonial feminist PAR is a new way of inquiry that

brings a number of discourses together to explore the concept of intersectionalities of influence in order to better understand the struggles and resiliences of immigrant and refugee women in the context of their health.

We have provided a brief introduction to postcolonial feminist theoretical perspectives and their relevance for nursing research and practice. While there continues to be debate about what constitutes postcolonial feminist scholarship, these theoretical perspectives help to generate transformative knowledge towards just, equitable, and improved care for immigrant and refugee women. We have argued for the relevance of the ecosystemic framework in operationalizing the concept of intersectionalities of influence in nursing research with immigrant and refugee women while locating their health concerns within the realm of family, community, and society. The framework has been useful in capturing the complexity of the theoretical domain of postcolonial feminism at the research practice level. We have introduced PAR as a research orientation that has been useful in our work with immigrant and refugee women and that fits well with the critical underpinnings of postcolonial feminism. Postcolonial feminist PAR incorporates knowledge, participation, and action, and brings issues of race, gender, class, culture, power, equity, and social justice to the forefront in research on health promotion among immigrant and refugee women. In combination, these approaches recognize women's agency and their positioning within hierarchical social spaces as a result of historical, socio-economic, cultural, and political contexts in both the country of origin and the host country. Such inquiry can benefit immigrant and refugee women while furthering the nursing research agenda.

References

Anderson, J. M. (2002). Toward a post-colonial feminist methodology in nursing research: Exploring the convergence of post-colonial and Black feminist scholarship. *Nurse Researcher, 9*(3), 7–27.

Anderson, J. M., Blue, C., Holbrook, A., & Ng, M. (1993). On chronic illness: Immigrant women in Canada's work force — A feminist perspective. *Canadian Journal of Nursing Research, 25*(2), 7–22.

Anderson, J., Perry, J., Blue, C., Browne, A., Henderson, A., Khan, K. B., et al. (2003). "Rewriting" cultural safety within the postcolonial and postnational feminist project: Toward new epistemologies of healing. *Advances in Nursing Science, 26*(3), 196–215.

Burns, N., & Grove, S. K. (2001). *The practice of nursing research: Conduct, critique, and utilization* (4th ed.). Philadelphia: W. B. Saunders.

Clough, P. T. (1992). *The end(s) of ethnography: From realism to social criticism*. Newbury Park, CA: Sage.

Clough, P. T. (1994). *Feminist thought: Desire, power and academic discourse*. Cambridge, MA: Blackwell.

Denton, M., Hadjukowski-Ahmed, M., O'Conner, M., & Zeytinoglu, I. U. (1999). A theoretical framework for research on women's health promotion. In M. Denton, M. Hadjukowki-Ahmed, M. O'Conner, & I.U. Zeytinoglu (Eds.), *Women's voices in health promotion* (pp. 9–20). Toronto: Canadian Scholars' Press.

Dhruvarajan, V., & Vickers, J. (2002). *Gender, race and nation: A global perspective*. Toronto: University of Toronto Press.

Duster, T. (2000). Foreword. In F.W. Twine & J.W. Warren (Eds.), *Racing research, researching race: Methodological dilemmas in critical race studies* (pp. xi–xiv). New York: NYU Press.

Gastaldo, D., Gooden, A., & Massaquoi, N. (2004). *Transnational health promotion: Social well being across borders and immigrant women's subjectivities.* Unpublished report.

Gastaldo, D., Khanlou, N., Massaquoi, N., & Curling, D. (2002). *Revisiting "personal is political": Immigrant women's health promotion*. Unpublished study funded by Canadian Institutes of Health Research.

Germain, C. B., & Bloom, M. (1999). *Human behaviour in the social environment*. New York: Columbia University Press.

Gilligan, C. (1982). *In a different voice*. Cambridge, MA: Harvard University Press.

Guruge, S. (2004). *Gender roles, power relations, and informal support systems in the context of post-migration domestic violence in the Sri Lankan Tamil Community of Toronto.* Doctoral dissertation proposal, Faculty of Nursing, University of Toronto.

Guruge, S., Lee, R. N. F., & Hagey, R. (2001). When conventional diabetes care is viewed as complementary: Perspectives of a Chinese client. *Alternative and Complementary Therapies, 7*(2), 117–120.

Hagey, R. (1997). Guest Editorial. The use and abuse of participatory action research. *Chronic Diseases in Canada, 18*(1), 1–4.

Hagey, R. (2000). Cultural safety: Honoring traditional ways of life. *Alternative and Complementary Therapies, 6*(4), 233–236.

Hagey, H., Choudhry, U., Guruge, S., Turrittin, J., Collins, E., & Lee, R. (2001). Immigrant nurses' experience of racism. *Journal of Nursing Scholarship, 33*(4), 389–394.

Harding, S. (1987). *Feminism and methodology.* Bloomington: Indiana University Press.

Harding, S. (1989). Is there a feminist method? In N. Tuana (Ed.), *Feminism and science* (pp. 17–32). Bloomington: Indiana University press.

Heise, L. L. (1998). Violence against women: An integrated, ecological framework. *Violence Against Women, 4*(3), 262–290.

Israel, B. A., Schulz, A. J., Parker, E. A., Becker, A. B., Allen, A. J. III, & Guzman, J. R. (2003). Critical issues in developing and following community based participatory research principles. In M. Minkler & N. Wallerstein (Eds.), *Community-based participatory research for health* (pp. 53–76). San Francisco: Jossey-Bass.

Khanlou, N. (2003). Mental health promotion education in multicultural settings. *Nurse Education Today, 23*(2), 96–103.

Khanlou, N., Beiser, M., Cole, E., Freire, M., Hyman, L., & Killbride, K. (2002). *Mental health promotion among newcomer female youth: Post-migration experiences and self-esteem*. Ottawa: Status of Women Canada.

Khanlou, N., & Hajdukowski-Ahmed, M. (1999). Adolescent self-concept and mental health promotion in a cross-cultural context. In M. Denton, M. Hajdukowski-Ahmed, M. O'Connor, & I. Zeytinoglu (Eds.), *Women's voices in health promotion* (pp. 138–151). Toronto: Canadian Scholars' Press.

Khanlou, N., & Peter, E. (in press). Participatory action research: Considerations for ethical review. *Social Science and Medicine*.

Kirsch, G. E. (1999). *Ethical dilemmas in feminist research: The politics of location, interpretation, and publication*. Albany: State University of New York Press.

Krug, E. G., Dahlberg, L. L., Mercy, J. A., Zwi, A. B., & Lozano, R. (2002). *World report on violence and health*. Geneva: World Health Organization.

McDonald, S. E. (2000). *The right to know: Women, ethnicity, violence and learning about the law*. Unpublished PhD dissertation, Ontario Institute of Studies in Education, University of Toronto.

Minkler, M., & Wallerstein, N. (2003). Introduction to community based participatory research. In M. Minkler & N. Wallersein (Eds.), *Community-based participatory research for health* (pp. 3–26). San Francisco: Jossey-Bass.

Peter, E. (2000). The politicization of ethical knowledge: Feminist ethics as a basis for home care nursing research. *Canadian Journal of Nursing Research, 32*(2), 103–118.

Racine, L. (2003). Implementing a postcolonial feminist perspective in nursing research related to non-Western populations. *Nursing Inquiry, 10*(2), 91–102.

Reimer-Kirkham, S., & Anderson, J. M. (2003). Postcolonial nursing scholarship: From epistemology to method. *Advances in Nursing Science, 25*(1), 1–17.

Seibold, C. (2000). Qualitative research from a feminist perspective in the postmodern era: Methodological, ethical and reflexive concerns. *Nursing Inquiry, 7*(3), 147–155.

Thorne, S., & Varcoe C. (1998). The tyranny of feminist methodology in women's health research. *Health Care for Women International, 19*(6), 481–493.

Tuhiwai Smith, L. (2001). *Decolonizing methodologies: Research and indigenous peoples*. Dunedin, NZ: University of Otago Press.

Turrittin, J., Hagey, R., Guruge, S., Collins, E., & Mitchell, M. (2000). The experiences of professional nurses who have migrated to Canada: Cosmopolitan citizenship or democratic racism? *International Journal of Nursing Studies, 39*, 655–667.

Varcoe C. (1996). Theorizing oppression: Implications for nursing research on violence against women. *Canadian Journal of Nursing Research, 28*(1), 61–78.

Varcoe, C. (2001). Abuse obscured: An ethnographic account of emergency nursing in relation to violence against women. *Canadian Journal of Nursing Research, 32*(4), 95–115.

Vickers, J. (2002). Methodologies for scholarship about women. In V. Dhruvarajan & J. Vickers, *Gender, race, and nation: A global perspective*. Toronto: University of Toronto Press.

Waller, M. A. (2001). Resilience in ecosystemic context: Evolution of the concept. *American Journal of Orthopsychiatry, 71*(3), 290–297.

Wilkinson, R., & Marmot, M. (1998). *Social determinants of health: The solid facts*. Copenhagen: World Health Organization.

Willinsky, C., & Pape, P. (1997). *Mental health promotion*. Toronto: Canadian Mental Health Association.

Authors' Note

The first author gratefully acknowledges support for her PhD studies in the form of a CIHR doctoral fellowship.

Comments or inquiries may be directed to Sepali Guruge, Faculty of Nursing, University of Toronto, 50 St. George Street, Toronto, Ontario M5S 3H4 Canada. E-mail: s.guruge@utoronto.ca

---

Sepali Guruge, RN, BScN, MSc, is a PhD candidate in the Faculty of Nursing, University of Toronto, Ontario, Canada. Nazilla Khanlou, RN, PhD, is Assistant Professor, Faculty of Nursing and Department of Psychiatry, University of Toronto.

Résumé

Les programmes de prévention de chutes chez les aînés : une méta-analyse secondaire bayesienne

Joseph F. Lucke

Une méta-analyse secondaire bayesienne des programmes pour réduire les chutes chez les aînés est effectuée pour démontrer le processus d'analyse bayesienne. La tradition statistique bayesienne fait l'objet d'une importante distinction relativement à la tradition statistique standard de Neyman-Pearson-Wald (NPW). Dans le cadre des 12 études, l'ampleur de l'effet logit est utilisée pour comparer des groupes thérapeutiques adhérant à un programme de prévention à des groupes témoins qui ne sont pas soumis à un programme. Afin de mettre en contraste l'analyse bayesienne, des méta-analyses d'effets indépendants et d'effets fixes sont d'abord réalisées selon la tradition de NPW. Cette procédure est suivie de méta-analyses d'effets indépendants et d'effets fixes qui reproduisent numériquement les résultats NPW mais qui comportent des interprétations différentes sur le plan conceptuel. Les dernières analyses comprennent des méta-analyses prédictives et des méta-analyses d'effets aléatoires bayesiennes. Ces résultats diffèrent sur le plan numérique de toutes les autres méta-analyses antérieures et sur le plan conceptuel des analyses de NPW. Les analyses des effets aléatoires permettent une hétérogénéité en ce qui a trait à l'ampleur de l'effet. L'analyse prédictive génère une distribution d'effet hors-échantillon nouveau, qui favorise non seulement l'hétérogénéité des effets mais aussi l'imprécision dans les estimations de paramètres. Cette dernière analyse démontre que l'efficacité des nouveaux programmes de prévention des chutes est moins définitive que celle relevée dans l'échantillon. Les méthodes statistiques bayesiennes se prêtent particulièrement bien au traitement des complexités contenues dans les recherches en sciences infirmières.

Mots clés : tradition statistique bayesienne, ampleur de l'effet logit, tradition statistique de Neyman-Pearson-Wald, analyse prédictive, méta-analyse secondaire

CJNR 2004, Vol. 36 N° 3, 48–64

Fall-Prevention Programs for the Elderly: A Bayesian Secondary Meta-analysis

Joseph F. Lucke

A secondary meta-analysis of programs to reduce falls in the elderly is undertaken to demonstrate a Bayesian analysis. The Bayesian statistical tradition is carefully distinguished from the standard Neyman-Pearson-Wald (NPW) statistical tradition. In the 12 studies, the logit effect size is used to compare treatment groups using a prevention program to control groups without a program. To contrast the Bayesian analysis, independent-effects and fixed-effect meta-analyses are first conducted in the NPW tradition. This is followed by Bayesian independent-effects and fixed-effect meta-analyses that numerically replicate the NPW results but have conceptually different interpretations. The final analyses comprise Bayesian random-effects and predictive meta-analyses. These results differ numerically from all the previous meta-analyses and conceptually from the NPW meta-analyses. The random-effects analysis allows for heterogeneity in the effect sizes. The predictive analysis yields the distribution of a new, out-of-sample effect size, which accommodates not only the heterogeneity of the effects but also the imprecision in the parameter estimates. This last analysis shows that the effectiveness of new fall-prevention programs is less definitive than that found in the sample. Bayesian statistical methods are particularly well-suited for the complexities of nursing science studies.

Keywords: Bayesian statistical tradition, fixed-effect model, health-care outcomes, hierarchical model, independent-effects model, logistic regression, logit effect size, Neyman-Pearson-Wald statistical tradition, predictive analysis, random-effects model, secondary meta-analysis

Hill-Westmoreland, Soeken, and Spellbring (2002) conducted a fixed-effect meta-analysis of the success of programs to prevent falls in the elderly. Their statistical analysis was based on the familiar Neyman-Pearson-Wald (NPW) statistical tradition. The purpose of this article is to illustrate the alternative Bayesian statistical tradition by conducting a secondary analysis of the same data. Bayesian statistical methods are rarely used in nursing science. A *fin de millénaire* review of research using Bayesian inference ranging from "archeology" to "social sciences" found no reference to nursing science (Berger, 2000). I will first present a brief sketch of the principles of the Bayesian statistical tradition.

The term "statistical tradition" is used here to emphasize the conceptual discontinuity between the two approaches. A *research tradition,* similar

to a paradigm (Kuhn, 1962), is a global cluster of beliefs about the entities and processes that make up the domain of inquiry and the methodological norms by which the domain is to be investigated (Laudan, 1981). A *statistical tradition* is by analogy a global set of ideas about the nature of probability and principles for statistical applications. Bayesian statistics is not another branch of NPW statistics but a comprehensive, alternative stance on probability and statistics. Use of the word "tradition" also emphasizes that concepts in the NPW tradition (e.g., significance, power) may have no meaning in the Bayesian tradition, that concepts in the Bayesian tradition (e.g., Bayes's factor, prior distribution) may have no meaning in the NPW tradition, and that identical numerical results may have radically different interpretations (e.g., confidence intervals, hypothesis tests). Table 1 provides a brief overview of the principal differences, for which only a brief comparison can be provided here.

The NPW statistical tradition interprets probability as relative frequency. Statistical inference is based on the fundamental concept of *inductive behaviour,* behaviour that can be repeatedly evaluated as correct or incorrect. The purpose of statistics is to develop rules for inductive behaviour and to assess their performance according to their relative frequency of success. With this approach, observations are considered random variables that are controlled by fixed but unknown parameters. For example, a statistical test will be acceptable if the rejection (behaviour) of a truly null (fixed and unknown) hypothesis will be incorrect only 5% of the time (performance) in the long run (probability). Confidence intervals, now favoured over significance tests (Altman, Machin, Bryant, & Gardner, 2000), provide the set of points that would not have been rejected by the corresponding significance test.

Table 1 ***Comparison of the Neyman-Pearson-Wald and Bayesian Traditions***

Topic	**NPW**	**Bayesian**
Probability	Relative frequency	Logic of judgement
Inference	Inductive behaviour	Degrees of belief
Statistic	Performance	Evidence
Observations	Random	Fixed
Parameters	Fixed	Random
Confidence interval	Random interval covers fixed parameter	Fixed interval contains random parameter

The Bayesian tradition interprets probability as a *logic of judgement* for those opinions that can be represented as *degrees of belief* (Howson & Urbach, 1993). This logic is regulated by *coherence,* a counterpart to consistency in deductive logic. Coherence ensures that a person cannot hold degrees of belief that are uniformly disadvantageous. The remarkable Ramsey-de Finetti Theorem shows that for judgements to be coherent the corresponding degrees of belief must satisfy the axioms of probability (Howson & Urbach). Thus, judgements can be represented by *subjective probabilities.* This approach considers observations to be fixed and the parameters of models that account for the data to be random. Bayesian confidence intervals or *credible intervals* give the fixed interval in which the random parameter can be found with the specified probability. It is often the case that NPW confidence intervals have the exact same numerical values as Bayesian credible intervals, but with different interpretations.

Because coherent judgements are subjective probabilities, the full use of the probability calculus is available for statistical inference. The Bayesian principle of inference is *Bayes's Theorem,* first posthumously published by Bayes and Price (1763) but a trivial theorem in modern probability theory. The application of Bayes's Theorem is a four-step process. First, the investigator develops the *likelihood,* which provides the probability of the observations y given the values of a parameter θ, written $\Pr(y \mid \theta)$. The likelihood provides the evidential link between the parameter and the observations (Royall, 1997). Second, because it is uncertain, the investigator represents the parameter θ as a random variable with a *prior* distribution $\Pr(\theta)$ that reflects uncertainty regarding its possible values. The prior distribution represents the information the investigator possesses before any observations are made in the current study. Third, the study is undertaken and the observations are made. Fourth, Bayes's Theorem provides the "updated" *posterior* distribution regarding the parameter conditioned on the observations. Bayes's Theorem states

$$\Pr(\theta \mid y) = \frac{\Pr(y \mid \theta)\Pr(\theta)}{\Pr(y)} . \qquad (1)$$

Roughly, Equation 1 states that

$$\text{Posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{data}} ,$$

or, sardonically, that "Bayesians can be recognized by their posteriors."

Bayesian statistical inference requires a prior distribution representing the investigator's information regarding the parameter of interest. The primary advantage of a prior is that it encapsulates previous knowledge regarding the parameter under investigation. The primary disadvantage is that an investigator's prior could also be quite idiosyncratic. However, several considerations moderate the choice of a prior. First, the chosen prior is exposed to the scrutiny of the scientific community. Determining the prior distribution forces the investigator to confront and make explicit his or her beliefs, both justifiable and speculative, regarding the phenomenon under investigation (Kadane, 1995). Second, the prior distribution is a *representation,* not a measurement, of one's degrees of belief (Hacking, 2001). Thus the investigator may *temper* the prior to more reasonably represent degrees of belief of the scientific community rather than actual beliefs (Shimony, 1970/1993). And third, a prior distribution can be made *diffuse* or *vague* so that it has little influence on the data (Edwards, Lindman, & Savage, 1963). The requirement of a publicly presented prior makes the subjective evaluation of prior knowledge more objective.

Meta-analysis is particularly well-suited for an introduction to Bayesian statistical analysis. The data sets tend to be small but fraught with inferential problems not found in usual analyses. Bayesian meta-analysis has a growing and increasingly sophisticated literature (Beard, Curry, Edwards, & Adams, 1997; Berlowitz et al., 2002; Brophy, Belisle, & Joseph, 2003; Brophy & Joseph, 2000; Brophy, Joseph, & Rouleau, 2001; Burr, Doss, Cooke, & Goldschmidt-Clermont, 2003; DuMouchel, 1990; Higgins & Spiegelhalter, 2002; Louis & Zelterman, 1994; Nam, Mengersen, & Garthwaite, 2003; Spiegelhalter & Best, 2003; Stangl & Berry, 2000; Warn, Thompson, & Spiegelhalter, 2002). The presentation here is intended to serve as an entry to Bayesian meta-analysis and to other Bayesian methods.

The remainder of this paper is organized as follows. First, the observed fall-prevention data and a metric for assessing treatment effects, namely the logit effect size, are presented. Second, two brief NPW analyses of the data, an independent-effects analysis and a fixed-effect analysis, are undertaken to provide a comparison for the Bayesian analyses. Third, two Bayesian analyses, also an independent-effects analysis and a fixed-effect analysis, are presented to compare to the NPW analyses. These two sets of analyses yield similar numerical results. The purpose is to demonstrate the conceptual differences between the two traditions even when the numbers are the same. And fourth, a random-effects Bayesian analysis coupled with a predictive analysis is presented. This last analysis shows how the Bayesian approach can utilize more realistic assumptions.

The Data and the Logit Effect Size

Table 2 presents the data on 12 select studies of falls in the elderly (Hill-Westmoreland et al., 2002). The first column contains the study number as enumerated by those authors. Each study i, $i = 1, \ldots, 12$, comprises two groups, a treatment group $j = 1$ which had an intervention for the prevention of falls, and a control group $j = 0$ which had no such intervention. Let y_{ij} denote the number of subjects experiencing one or more falls from a sample of n_{ij} subjects in group j of study i. Let $p_{ij} = y_{ij}/n_{ij}$ denote the proportion of falls. The *Control* columns give the observed y_{i0}, n_{i0}, and p_{i0}; the *Treatment* columns give the observed y_{i1}, n_{i1}, and p_{i1}.

Table 2 ***Meta-analytic Data on 12 Studies of Falls in the Elderly***

	Control		**Treatment**		**Comparison**	
Study	Falls/Sample	Prop.	Falls/Sample	Prop.	Logit	SE
1	62/117	0.53	53/116	0.46	-0.29	0.26
2	111/213	0.52	59/184	0.32	-0.84	0.21
3	45/94	0.48	34/75	0.45	-0.10	0.31
4	40/92	0.43	42/88	0.48	+0.17	0.30
5	129/261	0.49	91/221	0.41	-0.33	0.18
6	17/50	0.34	72/180	0.26	+0.40	0.34
7	61/81	0.75	56/79	0.71	-0.23	0.36
8	3/15	0.20	3/30	0.10	-0.81	0.89
9	6/13	0.46	2/14	0.14	-1.64	0.94
10	68/144	0.47	52/147	0.35	-0.49	0.24
11	8/47	0.17	68/332	0.20	+0.23	0.41
12	223/607	0.37	268/952	0.28	-0.39	0.11

Hill-Westmoreland et al. (2002) used the *difference effect*, $d_i = p_{i1} - p_{i0}$, to estimate the treatment effect (Rosenthal, 1994). Although the difference effect has a simple and straightforward interpretation, d_i has a number of defects, not least of which is that it is confounded with the baseline proportion found in the control group (Fleiss, 1994). A better measure of treatment effect, and the one used here, is the *logit effect*

$$w_i = \log\left(\frac{p_{i1}}{1 - p_{i1}}\right) - \log\left(\frac{p_{i0}}{1 - p_{i0}}\right) \qquad (2)$$

with standard error

$$SE(w_i) = \sqrt{\frac{1}{p_{i0}(1 - p_{i0})n_{i0}} + \frac{1}{p_{i1}(1 - p_{i1})n_{i1}}}$$

(Rosenthal). Although the logit is less intuitive than the difference, it is preferable on several grounds, two of which are given here. First, the logit effect is not confounded with the baseline proportion and may assume any value regardless of the proportions being compared (Fleiss). And second, the logit effect is a natural parameter of log-linear and logistic regression models (Fleiss). The logit effects and their standard errors for the 12 studies are also given in the last two columns of Table 2. Negative logits favour treatment over control.

Neyman-Pearson-Wald Analyses

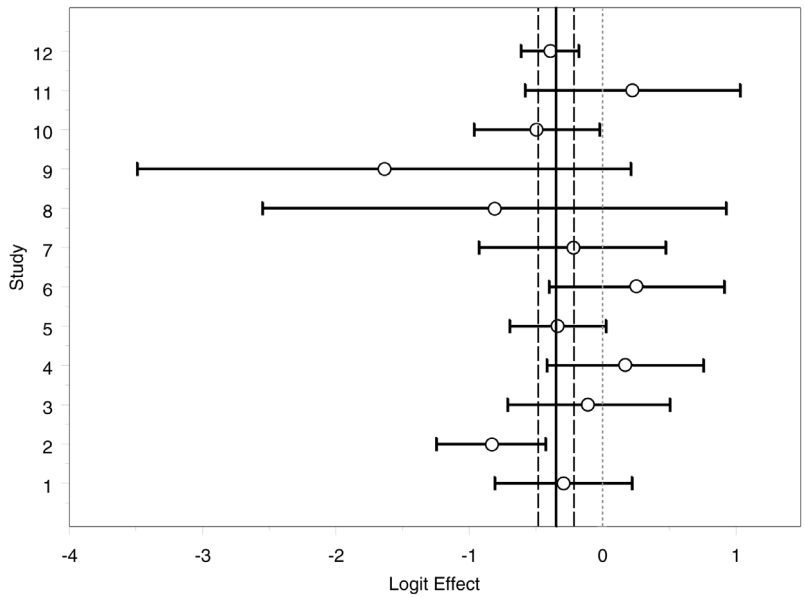
Independent-Effects Analysis

The independent-effects analysis addresses the treatment effect of each study separately from any other study. It is not a meta-analysis itself, but such analyses are invariably included in a meta-analysis to display the individual effects. The sample logit effects follow an asymptotic normal distribution with mean w_i and standard deviation $SE(w_i)$. Figure 1 displays the logit effect for each study together with the respective 95% confidence intervals. Only three studies, 2, 10, and 12, had 95% confidence intervals that excluded zero, or, equivalently, rejected the null hypothesis of $w_i = 0$ with a two-tailed significance at .05. Hill-Westmoreland et al. (2002), using difference effects, found Study 9 to be significant and not Study 10. Thus, the difference effect and logit effect need not produce identical results.

Fixed-Effect Analysis

A fixed-effect meta-analysis assumes the existence of a fixed but unknown population treatment effect. Each study provides a sample estimate of the population effect. The estimator of the population fixed effect is the weighted mean of the study sample effects and likewise follows an asymptotic normal distribution (Fleiss, 1994). The first row of Table 3 presents the weighted mean logit effect of $w = -0.35$, its standard error, and its 95% confidence interval. Because the interval for this statistic excludes 0, the null hypothesis of no overall treatment effect is rejected. This fixed-effect re-analysis using logit effects yields the same conclusion as Hill-Westmoreland et al.'s (2002) original analysis.

Figure 1 ***Forest Plot of Logit Effect Sizes for the NPW Analyses***



Note: Circles with horizontal bars represent independent logit effects and their respective 95% confidence intervals. The solid vertical line is the fixed logit effect and the dashed vertical lines are its 95% confidence interval.

Table 3 ***Meta-analytic Results from NPW and Bayesian Analyses***

Analysis		Logit Effect	Standard Error	Confidence Interval	Hypothesis: Effect < 0
NPW	Fixed	-0.35	0.07	-0.48, -0.21	$z = -5.07$
Bayesian	Fixed	-0.35	0.07	-0.48, -0.21	$\Pr(\beta < 0) > 0.999$
Bayesian	Random	-0.32	0.10	-0.51, -0.11	$\Pr(\mu_\beta < 0) = 0.996$
Bayesian	Predictive	-0.32	0.23	-0.79, +0.20	$\Pr(B < 0) = 0.93$

Bayesian Analyses I

Independent-Effects Analysis

A Bayesian independent-effects analysis is presented here for comparison with the NPW version. Again, this analysis is, strictly speaking, not a meta-analysis, and is not even needed as an intermediary step in the

Bayesian analysis. Nevertheless, it is frequently presented to display the individual effects. The likelihood is based on the binomial-logistic regression model (McCaullagh & Nelder, 1989). The propensity of a subject to experience at least one fall in the control group of study i is the logistic function of α_i. The propensity to fall in the treatment group is the logistic function of $\alpha_i + \beta_i$, where β_i is the treatment effect. If the treatment reduces the propensity to fall, β_i will be negative. The Bayesian analysis requires prior distributions for the parameters α_i and β_i. Because there is no prior information on these parameters and because the desire here is for the observations to have maximal influence on the posterior distributions, diffuse priors were chosen for both sets of parameters. In particular, the priors were chosen to be normal distributions with mean of zero and standard deviation of 1,000. These priors stipulate that this investigator was 95% sure that the parameter values lay between -2,000 and 2,000.

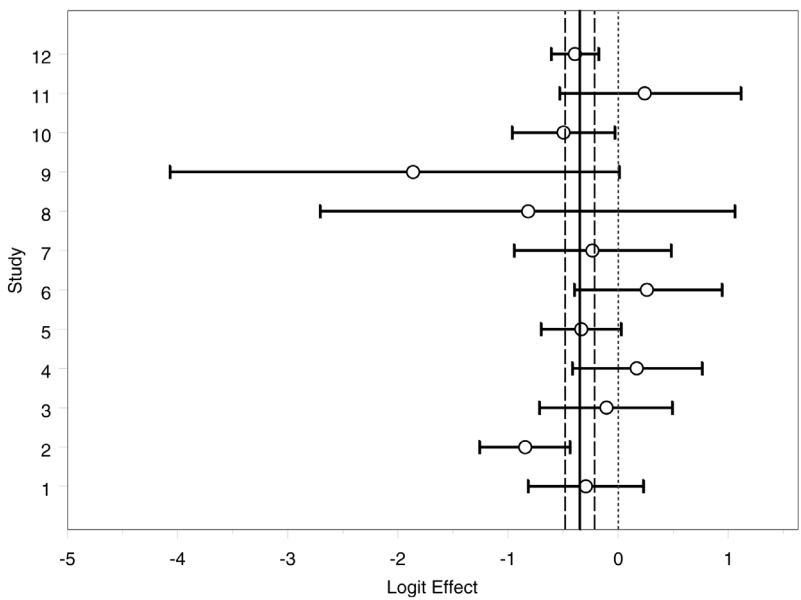
Equation 1 in principle supplies the posterior distributions of the 12 α_i's and the 12 β_i's. However, substituting the appropriate formulae from the distributional assumptions for the likelihood of γ_{ij} and the priors for the α_i and the β_i yields a formula that is analytically intractable. Numerical methods must be employed to obtain the posterior distributions. One popular approach is to numerically simulate the posterior distributions of each parameter to a high degree of accuracy. Given the simulated posteriors, summary information regarding the parameters can be readily obtained. The current widely used procedure for obtaining numerical solutions is *Markov Chain Monte Carlo* (MCMC) simulation using Gibbs sampling (Gelman, Carlin, Stern, & Rubin, 2004). MCMC is an algorithm in which samples are taken from each parameter's distribution in a round-robin fashion and used to update the other distributions in the model. This sampling procedure is iterated several thousand times, and under suitable mathematical conditions the distributions of the parameters will converge to a unique set of limit distributions. These limit distributions can then be used to extract statistical information regarding the parameters. This independent-effects model was programmed in WinBUGS 1.4 (Spiegelhalter, Thomas, Best, & Lunn, 2002). To stabilize the starting distributions, an initial (burn-in) run of 2,000 iterations was taken without sampling the parameters. The sampling simulation was then run for another 20,000 iterations to obtain the limit distributions. The simulation takes about 12 seconds on a reasonably fast (1 GHz) desktop computer. S-PLUS Professional 6.2 (Insightful Corporation, 2001) and Microsoft Excel were used for additional computations and graphics.

Figure 2 displays the independent effects together with their 95% Bayesian credible intervals. These are very similar in numerical value to

the 95% confidence intervals of the NPW analysis in Figure 1. The single exception appears to be study 9, which has a smaller Bayesian posterior mean and larger standard error $\beta_9 = -1.86$, se = 1.04 than the corresponding NPW estimate of $w_9 = -1.64$, se = 0.94. This discrepancy is most likely due to the inaccuracy of the NPW asymptotic estimator in a small sample of $n_{9,0} + n_{9,1} = 27$ subjects.

More important than the numerical similarity between the NPW and Bayesian results is that the interpretation is completely different. The credible intervals give the fixed interval in which the random treatment effect probably lies, in sharp contrast to the NPW interpretation of the random interval that probably covers the fixed treatment effect.

Figure 2 ***Forest Plot of Logit Effect Sizes for the Bayesian Analyses***



Note: Circles with horizontal bars represent independent logit effects and their respective 95% credible intervals. The solid vertical line is the fixed logit effect and the dashed vertical lines are its 95% credible interval.

Fixed-Effect Analysis

The fixed-effect analysis again assumes that there is a unique treatment effect for all the studies and that each study displays this unique effect accompanied by random error. The fixed-effect analysis merely replaces each study-specific β_i in the independent-effects analysis with a single β for all studies in the binomial-logistic regression. Thus the treatment

effect is, instead, a logistic function of $\alpha_i + \beta$. The prior distributions for the 12 α_i's and the one β are the same normal distributions as in the independent-effects case.

Again, the resulting posterior distribution is analytically intractable and MCMC simulation was used. The second row of Table 3 presents the Bayesian fixed-effect result. The numerical values are identical to those of the NPW analysis but the interpretation in the Bayesian case is that $\Pr(-0.48 \leq \beta \leq -0.21) = .95$. Furthermore, because β is a random variable, probabilistic hypotheses can be entertained. Of interest here is whether the fall-prevention programs reduced the proportion of falls, which translates into the hypothesis whether β is negative. The probability is that $\beta < 0$ can be readily evaluated from its relevant posterior distributions. The rightmost column of Table 3 shows that $\Pr(\beta < 0) > .999$.

Bayesian Analyses II

Random-Effects Analysis

The purpose of the Bayesian fixed-effect analysis was to demonstrate that the Bayesian approach can replicate the numerical results of the NPW approach, even though the conceptual interpretations of the results are different and incompatible between the two traditions. The purpose of the random-effects analysis is to show how the Bayesian approach can easily accommodate more realistic assumptions in the statistical model.

The assumption of a unique, fixed treatment effect for each study is excessively restrictive. It implies that all differences among study effects are due solely to random error. It ignores possible differences in settings and implementations of the prevention programs that would have caused different effects in addition to random fluctuations. A more realistic and plausible assumption is that the population effects themselves arise from a distribution of treatment effects. The crucial assumption required for such an analysis, called *exchangeability*, is that the magnitudes of the program effects are equally as likely to appear in one study as in another (Gelman et al., 2004). The assumption of exchangeability allows for heterogeneous but related effects in place of the assumption of a homogeneous effect.

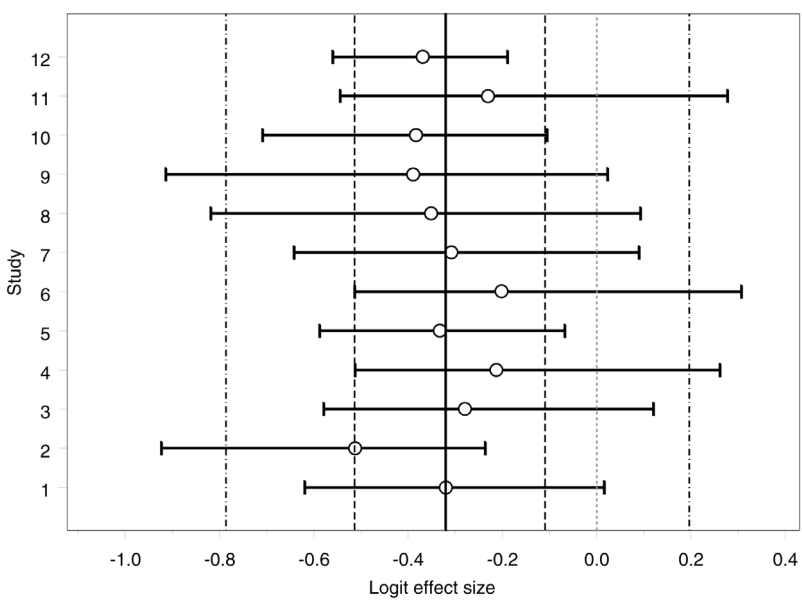
As in the previous analyses, the binomial-logistic regression model is assumed. The baseline propensity to experience a fall is assumed to be specific to each study, implying that the α_i's are again assumed to be fixed. The β_i's, however, are not assumed to be fixed for each study but assumed to be part of an overall effect of fall prevention on all the studies. To accommodate heterogeneity of effects among the studies, the β_i's are themselves assumed to be independently distributed random variables from a normal distribution with unknown mean μ_β and unknown

variance σ^2_β. This likelihood is equivalent to a two-level logistic regression (Rice, 2001).

To complete the Bayesian specification, prior distributions are required for the 12 α_i's, μ_β, and σ^2_β. Again, diffuse priors are chosen so that prior beliefs have little influence on the observed data. The fixed α_i's are once again given a normal prior with a mean of zero and a standard deviation of 1,000. The priors for the parameters of the distribution of the random treatment effects are similarly diffuse. The prior for μ_β is a normal distribution with a mean of zero and a standard deviation of 1,000, and the prior for σ^2_β is a diffuse gamma distribution. Again, the formula for the posterior distribution is analytically intractable, so that MCMC methods are used to simulate the posterior distributions.

Figure 3 presents the posterior mean logit effect for each of the 12 studies together with their 95% credible intervals. The first feature to note is that all the logit effects are negative, in contrast to the independent-effects analysis in which the effects of studies 4, 6, and 11 were positive.

Figure 3 ***Forest Plot of Random Logit Effect Sizes***



Note: Circles with horizontal bars denote the posterior mean logit effect sizes and their respective 95% credible intervals. The solid vertical line is the posterior mean logit effect size for the distribution of effects. The dashed vertical lines contain its 95% credible interval. The dot-dashed vertical lines contain the 95% credible interval for the predictive posterior mean effect size.

The second feature to note, related to the first, is that the logit effects tend to be closer together than in the independent-effects analysis, with no extreme cases. (The x-axis ranges only from -1 to 0.4 rather than from -5 to 2 for the independent-effects analysis.) These two features demonstrate the well-known phenomena of "shrinkage towards the mean" and "studies borrowing strength from each other" found in random-effects analyses.

The mean posterior mean treatment effect is $\mu_\beta = -0.32$ with $Pr(-0.51 \leq \mu_\beta \leq -0.11) = .95$. This mean and credible interval is also displayed in Figure 3. As shown in Table 3, the mean posterior mean and its standard error are slightly larger than that of the fixed logit effect. This shrinkage towards zero accompanied by a larger standard error is also common in random-effects models. The probability that μ_β is negative is .996.

Predictive Analysis

The random-effects analysis also yields the posterior distribution of the logit treatment effects. The distribution of treatment effects is a normal distribution (by assumption) with a posterior mean for $\mu_\beta = -0.32$ and a posterior mean for $\sigma_\beta = 0.17$. Thus it would be reasonable to predict that any treatment effect not yet observed would on average be -0.32 with a 95% credible interval of [-0.67, + 0.02]. However, a logical defect in making this particular out-of-sample prediction is that only the means of the distributions of μ_β and σ_β are used. A more realistic out-of-sample prediction would include the imprecision of μ_β and σ_β as well. Let *B* denote the out-of-sample treatment effect. A Bayesian *predictive analysis* obtains the distribution of *B* by sampling it from a normal distribution with mean μ_β and variance σ^2_β. However, μ_β and σ_β are not fixed points but random variables. Thus, the posterior distribution of *B* will have the same mean as the posterior distribution of μ_β but will have a larger variance than σ^2_β, comprising the variance of the normal distribution, the posterior variance of μ_β, and the posterior variance of σ^2_β.

Figure 3 presents the posterior predictive mean $B = -0.32$ with a 95% credible interval ranging from -0.79 to +0.20, considerably wider than the point-based credible interval mentioned in the previous paragraph. The hypothesis test shows that *B* only has probability of .93 of being negative.

Discussion

The Bayesian statistical tradition offers a comprehensive view of the nature of probability and statistical inference. Probability is interpreted as a logic of judgement regarding degrees of belief, and statistical inference is the revision of those subjective judgements according to observational

evidence. Unobserved quantities are random variables, known observations are fixed, inferences are based only on the observations, hypotheses can be probabilistically compared one with another, and credible intervals give the interval that probably contains the unobserved quantity. The NPW tradition also offers a comprehensive view of the nature of probability and statistical inference. Probability is interpreted as relative frequency, and statistical inference is the evaluation of the performance of decision rules for inductive behaviour. Unknown parameters are fixed, known observations are random, inferences are not based solely on the observations, statistical hypotheses are choices between inductive behaviours, and $1 - \alpha$-level confidence intervals give the interval that contains the point-null hypotheses that would not have been rejected by an α-level statistical test. Comparisons between the two traditions cannot be made on statistical grounds because each tradition contains its own, internal standards of evaluation, standards that need not be applicable to the other tradition. Choosing between traditions must instead be based on higher-order cognitive values regarding the goals and norms of scientific inference (Laudan, 1984). Many statisticians and philosophers of science believe that the Bayesian tradition offers a natural and commonsense interpretation of probability and statistical inference. These same authors often remark on the persistent tendencies of investigators to interpret *p* values as evidential support and confidence intervals as fixed intervals of probability as indications of the inferential perversity of the NPW tradition.

The demonstration here shows that a Bayesian statistical analysis can replicate an NPW analysis by yielding the same numerical results. The fixed-effect meta-analyses addressed whether the fall-prevention programs reduced the proportion of falls — that is, had an effect in the negative direction. The NPW approach yielded a logit effect size of $w = -0.35$ with a 95% confidence interval ranging from -0.48 to -0.21. The Bayesian approach likewise yielded a logit effect size of $\beta = -0.35$ with a 95% credible interval ranging from -0.48 to -0.21. These two identical numerical results have incommensurable interpretations. The NPW confidence interval holds that the set of point-null hypotheses from -0.48 to -0.21 would not have been rejected by a two-tailed test at the 5% significance level. The Bayesian credible interval holds that β falls between -0.48 and -0.21 with probability .95.

The random-effects analysis further showed how a Bayesian analysis could go beyond the assumption of a fixed effect by allowing a distribution of effects. By assuming a distribution of effects, the Bayesian analysis accommodated possible heterogeneity among study-specific effects and yielded the overall mean effect. The Bayesian analysis also yielded the entire distribution of effects along with its mean $\mu_\beta = -0.32$ and standard

deviation σ_β= -0.17. The Bayesian predictive analysis also yielded the distribution of an out-of-sample effect that included not only the variability in effects but also the variability in estimating the distribution of the effects. This final result indicates that future fall-prevention programs can expect, with 95% confidence, to have a logit effect size between -0.8 and +0.2. Thus the predictive analysis indicates that the effect of fall-prevention programs is less definitive than might be presupposed from the either the fixed-effect or random-effects analyses.

To begin using Bayesian methods, one must be prepared to undertake substantial changes in one's ideas about statistics. This demonstration, though simple, was sufficient to uncover some of the advantages of Bayesian statistical methods. The Bayesian approach is known for its ability to handle complicated and unusual situations (Best, Spiegelhalter, Thomas, & Brayne, 1996). Thus, the Bayesian statistical tradition appears well-suited for the difficult inferential problems found in nursing science.

References

Altman, D. G., Machin, D., Bryant, T. N., & Gardner, M. J. (Eds.). (2000). *Statistics with confidence: Confidence intervals and statistical guidelines* (2nd ed.). London: BMJ Books.

Bayes, T., & Price, R. (1763). An essay toward solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions, 53,* 370–418.

Beard, M. T., Curry, E. L., Edwards, K., & Adams, B. N. (1997). Advances in meta-analysis as a research method. *ABNF Journal, 8*(5), 92–97.

Berger, J. O. (2000). Bayesian analysis: A look at today and thoughts of tomorrow. *Journal of the American Statistical Association, 95*(452), 1269–1276.

Berlowitz, D. R., Christiansen, C. L., Brandeis, G. H., Ash, A. S., Kader, B., Morris, J. N., & Moskowitz, M. A. (2002). Profiling nursing homes using Bayesian hierarchical modeling. *Journal of the American Geriatrics Society, 50*(6), 1126–1130.

Best, N. G., Spiegelhalter, D. J., Thomas, A., & Brayne, C. E. G. (1996). Bayesian analysis of realistically complex models. *Journal of the Royal Statistical Society. Series A (Statistics in Society), 159*(2), 323–342.

Brophy, J. M., Belisle, P., & Joseph, L. (2003). Evidence for use of coronary stents: A hierarchical Bayesian meta-analysis. *Annals of Internal Medicine, 138*(10), 777–786.

Brophy, J. M., & Joseph, L. (2000). A Bayesian meta-analysis of randomized megtrials for the choice of thrombolytic agents in acute myocardial infarction. In D. K. Stangl & D. A. Berry (Eds.), *Meta-analysis in medicine and health policy.* New York: Marcel Dekker.

Brophy, J. M., Joseph, L., & Rouleau, J. L. (2001). β-blockers in congestive heart failure: A Bayesian meta-analysis. *Annals of Internal Medicine, 134*(7), 550–560.

Burr, D., Doss, H., Cooke, G. E., & Goldschmidt-Clermont, P. J. (2003). A meta-analysis of studies on the association of the platelet p1a polymorphism of glycoprotein iiia and risk of coronary heart disease. *Statistics in Medicine, 22,* 1741–1760.

DuMouchel, W. (1990). Bayesian metaanalysis. In D. A. Berry (Ed.), *Statistical methodology in the pharmaceutical sciences* (pp. 509–529). New York: Marcel Dekker.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70,* 193–242.

Fleiss, J. F. (1994). Measures of effect size for categorical data. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 245–260). New York: Russell Sage Foundation.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). London: Chapman & Hall.

Hacking, I. (2001). *An introduction to probability and inductive logic.* Cambridge: Cambridge University Press.

Higgins, J. P., & Spiegelhalter, D. J. (2002). Being sceptical about meta-analyses: A Bayesian perspective on magnesium trials in myocardial infarction. *International Journal of Epidemiology, 31*(1), 96–104.

Hill-Westmoreland, E. E., Soeken, K., & Spellbring, A. M. (2002). A meta-analysis of fall prevention programs for the elderly: How effective are they? *Nursing Research, 51*(1), 1–8.

Howson, C., & Urbach, P. (1993). *Scientific reasoning: The Bayesian approach* (2nd ed.). Chicago: Open Court.

Insightful Corporation. (2001). *S-PLUS 6 for Windows user's guide.* Seattle: Author.

Kadane, J. B. (1995). Prime time for Bayes. *Controlled Clinical Trials, 16,* 313–318.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Laudan, L. (1981). A problem-solving approach to scientific progress. In I. Hacking (Ed.), *Scientific revolutions* (pp. 144–155). Oxford: Oxford University Press.

Laudan, L. (1984). *Science and values: The aims of science and their role in scientific debate.* Berkeley: University of California Press.

Louis, T. A., & Zelterman, D. (1994). Bayesian approaches to research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 411–422). New York: Russell Sage Foundation.

McCaullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

Nam, I.-S., Mengersen, K., & Garthwaite, P. (2003). Multivariate meta-analysis. *Statistics in Medicine, 22,* 2309–2333.

Rice, N. (2001). Binomial regression. In A. H. Leyland & H. Goldstein (Eds.), *Multilevel modelling of health statistics.* Chichester: John Wiley.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.

Royall, R. M. (1997). *Statistical evidence: A likelihood paradigm.* London: Chapman & Hall.

Shimony, A. (1993). Scientific inference. In *Search for a naturalistic world view: Scientific method and epistemology* (Vol. 1). Cambridge: Cambridge University Press. (Original work published 1970)

Spiegelhalter, D. J., & Best, N. G. (2003). Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-effectiveness modelling. *Statistics in Medicine, 22,* 3687–3709.

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Lunn, D. (2002). *WinBUGS Version 1.4 user's manual.* MRC Biostatistics Unit. Available: http://www.mrc-bsu.cam.ac.uk/bugs/

Stangl, D. K., & Berry, D. A. (2000). *Meta-analysis in medicine and health policy.* New York: Marcel Dekker.

Warn, D. E., Thompson, S. G., & Spiegelhalter, D. J. (2002). Bayesian random effects meta-analysis of trials with binary outcomes: Methods for the absolute risk difference and relative risk scales. *Statistics in Medicine, 21,* 1601–1623.

Author's Note

I thank Dr. Hill-Westmoreland for generously providing the data.

Joseph F. Lucke, PhD, is Assistant Professor and Statistical Scientist, School of Nursing, University of Texas Health Science Center at San Antonio, USA.

Résumé

La technique fondée sur une échelle d'évaluation d'un produit permettant de mieux évaluer la satisfaction d'un patient

Mary R. Lynn et Bradley J. McMillen

Dans le but de trouver une solution à la pénurie actuelle d'infirmières dans les communautés autochtones du Nord-Ouest de l'Ontario, la Direction générale de la protection de la santé des Premières nations et des Inuits à Santé Canada a commandé une étude sur la pertinence d'établir une équipe de relève constituée d'infirmières provenant des petites villes minières des environs. On a présenté un questionnaire à questions libres et à questions fermées à un échantillon aléatoire de 237 infirmières en vue d'analyser leur degré de sensibilisation, leur disposition et leur niveau de préparation à la pratique des soins infirmiers en région nordique et de déterminer quels sont les facteurs favorables et défavorables au recrutement. Les conclusions révèlent une connaissance du recoupement des dimensions professionnelles et personnelles qui caractérise cette pratique; elles justifient le bien-fondé d'un système de rotation qui chevaucheraient les compétences fédérale, provinciale et locale. Malgré sa complexité, avec du temps et de la volonté, ce type de structure de relève régionale semble viable.

Mots clés : technique fondée sur une échelle d'évaluation d'un produit, satisfaction du patient, mesurage, outil

CJNR 2004, Vol. 36 N° 3, 66–81

The Scale Product Technique as a Means of Enhancing the Measurement of Patient Satisfaction

Mary R. Lynn and Bradley J. McMillen

Measurement of patient satisfaction has long been hampered by two resolvable problems: a lack of content validity in commonly used instruments, and a lack of variability in satisfaction scores when these same instruments are used. Most patient satisfaction instruments have been developed from the perspective of the provider or institution rather than that of the patient, creating a situation of questionable content validity for these measures. Additionally, most patient satisfaction measures yield data that are invariant and consistently positively biased. Both of these problems can be addressed methodologically — through tool development using a qualitative method designed to obtain the patient's perspective, and through the use of the scale product technique to decrease the effect of acquiescence, thereby increasing variability in item responses.

Keywords: measurement, Likert, scale product, patient satisfaction, patient-centred

Patient satisfaction surveys are commonly used to evaluate health care, primarily because the patient's input into the evaluation of care is almost universally seen as essential (Committee on Quality Health Care in America, 2001; Donabedian, 1980; Walker, 1993). In several models of health care, patient satisfaction is considered *the* outcome of care (Oberst, 1984), or one of two or three outcomes of care (Ellencweig, 1992; Jelinek, 1967). Elling (1974, 1980) suggests that patient satisfaction with care is one of five secondary objectives that are antecedent to the primary objective of desirable health status.

However, as important as patient satisfaction is in the evaluation of care, the current state of its measurement is severely lacking, for two reasons. First, despite recent advances in the assessment of satisfaction among hospitalized patients (e.g., Yen, Chen, & Chou, 2003), many items commonly used in patient satisfaction tools still tend to address either "hotel services" (Louden, 1989) or the perspective of the provider rather than that of the patient (Corriher, 1994; McDaniels & Nash, 1990; Rubin, 1990). And second, these items (or any others), once placed into a instrument with the now standard five-point Likert balanced response format (strongly disagree to strongly agree), result in a positively biased or a highly acquiescent response set (Ware, Davies-Avery, & Stewart, 1977). Since it is difficult to disagree with items that vary minimally (e.g.,

"the nurses acted friendly") or have little relevance for the evaluation of care (e.g., "my food was served hot"), it is not surprising to find that patient satisfaction tools generally yield minimally useful data (McDaniels & Nash).

However, both of these restrictions on the utility of patient satisfaction measures can be addressed methodologically. First, despite the progress that has been made in this area (e.g., Hancock et al., 2003; Hilton, Budgen, Molzahn, & Attridge, 2001; Yen, Chen, & Chou, 2003), qualitative methods should still be used to determine what patients value about nursing care and what they expect when hospitalized. Only with such approaches can the validity of patient satisfaction scales be maximized. Second, the use of a response format that includes a method for enhancing variability and decreasing the positive response bias commonly found will advance the measurement of patient satisfaction. This paper reports on a method for achieving such enhancements by using items derived from qualitative interviews conducted with patients to determine what they consider important in nursing care received in the acute-care setting, combined with a method to augment traditional scaling of such items.

The Patient's Perspective in Quality Nursing Care

Uncertainty about what patients consider important in nursing care has been mentioned in a number of discussions on patient satisfaction (Bond & Thomas, 1991; Murdaugh, 1992; Oberst, 1984; Ware et al., 1977). Substantial qualitative work has clearly demonstrated that patients can and do articulate the specific aspects of care they consider important in nursing care (Fosbinder, 1994; Lynn & Sidani, 1995). In fact Lynn and Sidani (1995) identify 90 specific aspects of care that patients consider the essential components of quality nursing care. These include having a nurse who is able to determine, based on her/his knowledge, what the patient needs; instructs the patient in a clear and understandable manner; calls the patient by his/her preferred name; and encourages the patient to become involved in her/his own care.

However, a patient satisfaction tool based on these 90 aspects will be sufficiently sensitive to the subtleties in individual patient evaluations only if it considers the relative importance of each aspect. Traditional Likert scaling, the most common scaling method chosen for most tools, including patient satisfaction ones, does not allow for items with different degrees of importance; in fact, it assumes that all items are of equal intensity or strength (Likert, 1932; van Alphen, Halfens, Hasman, & Imbos, 1994). If Likert items could have varying degrees of importance, and if patient satisfaction tools incorporated this revised scoring format,

then perhaps the response bias limitation of the patient satisfaction scales could be overcome.

Oberst (1984) poses the central question regarding varying levels of patient satisfaction with nursing care: "Is there a hierarchy of satisfaction that can be identified and are certain aspects of [care] more satisfying?" (p. 2367). This hierarchy question can be answered methodologically using the judgement technique developed by Thurstone (1928). Once the judgement stage is reached, the judgements can be directly converted to weights for use in combination with Likert scaling (Likert, 1932) to enhance the sensitivity of the patient satisfaction measurement. This combination of Thurstone judging and Likert scaling is the scale product technique developed by Eysenck and Crown (1949) and used by Hulka, Zyzanski, Cassel, and Thompson (1970) and Zyzanski, Hulka, and Cassel (1974) in studies of patient evaluation of medical care.

Thurstone or Likert Scaling — or Both?

Among the many decisions made when an instrument is developed, the choice of scaling method is particularly important. The scaling method developed by Likert (1932), a summated rating method, is one of the most common methods used by those developing or revising instruments. With the Likert scale, the subject is placed on a continuum according to her/his degree of agreement or disagreement on the topic being measured (McIver & Carmines, 1981). This is done by summing the subject's responses using a balanced continuous response format (usually ranging from strongly disagree to strongly agree) for each item in the scale to derive a total score, which is assumed to represent the person's relative position on the topic. Likert developed his method of scaling in response to the extensive effort entailed in the dominant scaling technique of the time, the equal appearing interval scaling method developed by Thurstone (Thurstone, 1928; Thurstone & Chave, 1929). Perhaps a better scaling method is one that combines the best of these two approaches, the scale product method (Eysenck & Crown, 1949; Eysenck, Crown, & Shapiro, 1950). After a review of the Thurstone and Likert methods, the scale product method will be discussed, using an example, as a means of addressing the shortcomings of the two classic methods.

Thurstone Equal Appearing Interval Scales

L. L. Thurstone, a pioneer in the development of attitudinal measurement, conceived attitude measurement as the judgement concerning one stimuli in reference to or compared to another. In his method of equal appearing intervals, the researcher generates a large pool of statements

pertaining to the attribute to be measured. Ideally, these are simple, declarative statements that a person can either agree or disagree with, and represent a range of intensity or strength of opinion regarding the concept being measured.

The statements are then written on cards and given to a group of individuals who serve as judges. The judges rate each statement according to its placement along a continuum, reflecting the extent to which it represents a positive or negative attitude towards the concept being measured. Items are typically rated on an 11-interval scale, from A = unfavourable to K = favourable. Only the middle (neutral) and the two extremes are defined for the judges (Edwards, 1957). After each item is rated by the judges, its median ranking is determined using the centile formula to obtain the 50th percentile, which becomes the scale score associated with that item (Stevens, 1946). Only the items that have stable median rankings (those with relatively low semi-interquartile ranges) are retained for use in the scale (Edwards). The desired outcome of this process is a scale with approximately two items for each of the 11 equal appearing intervals. When there are more than two items per interval, the items chosen are those with the lowest semi-interquartile ranges.

The "correct" number of judges to employ in the Thurstone method depends on the number needed to ensure substantial agreement on the placement of items along the continuum. Thurstone's original study used 300 judges, although subsequent research using far fewer judges has produced sufficient agreement on item placement along the continuum (Crocker & Algina, 1986). When a Thurstone scale is administered, respondents indicate whether they agree or disagree with each item. The average scale score (0–10) of the items that the person agrees with becomes his/her score.

One of the drawbacks of the Thurstone method is that it is quite laborious (Likert, 1932; Murphy & Likert, 1938). A substantial amount of time is required to find the judges and to generate the dozens of items needed to ensure an 11-point range. Also, it can be difficult to find items that fall into the moderate intervals of the scale (Edwards, 1957; Ferguson, 1939) and to secure agreement on the scale value of an item, which, by definition, falls into the grey area between quite favourable and quite unfavourable. The Thurstone method is also criticized because it assumes that the concept being measured is unidimensional, which is not a reasonable assumption for many affective phenomena.

Likert Summated Rating Scales

The Likert method of "summated ratings" came about because of the logistical difficulties of the Thurstone method. The items in a Likert scale are split between those that are positively worded (favourable) and those

that are negatively worded (unfavourable). Unlike items in the Thurstone method, those in the Likert scale do not have to be distributed across a continuum; in fact, all of the items are assumed to be of the same magnitude in favouring (or not favouring) the measurement objective of the instrument. It is this last assumption, that all items are of approximately the same strength or intensity, that mathematically allows for the items to be summed, with equal weighting, to derive the total score. Interestingly, Likert items, when rated according to the Thurstone method, tend to cluster at the top and bottom of the 11-point continuum (Ferguson, 1939).

The Likert scale has a more precise response format than the Thurstone scale. Its typical response format has five options, ranging from strongly disapprove/disagree to strongly approve/agree, with an undecided/neutral central point (Likert, 1932). Scores are obtained by assigning integer values to item responses and summing them to derive a total score (e.g., strongly disagree = 1, disagree = 2, neither clearly agree nor clearly disagree = 3, agree = 4, strongly agree = 5). It should be noted, however, that a Likert response format can have from 3 to 21 response categories (although 3 to 7 is the usual range), and a neutral response option is not a requirement. Many different labels have been used for Likert response formats; no one response format is required in a Likert scale; the only requirement is that the response options be balanced, with equal numbers of agree and disagree options.

The Likert method, however, also has its shortcomings. All items count equally in the total score. Therefore, agreeing (or disagreeing) with any one item contributes just as much to a person's score as agreeing (or disagreeing) with any other item. As stated above, the equal value of each item in the total score is based on the assumption that each item is as important as the next in terms of its centrality to the concept being measured. Therefore, a Likert scale lacks the scale values that are generated by Thurstone judging, which can provide valuable information about the items.

Another limitation of the Likert method is what can be called the "ocular test" of item quality. While the Thurstone method allows one to weed out items of suspect quality by eliminating those that receive ambiguous rankings during the judging, the Likert method does not allow for direct assessment of item quality.

Scale Product Technique

In light of these shortcomings, the Thurstone and Likert methods were combined for the purpose of creating the scale product technique, which capitalizes on the strengths of both methods while ameliorating some of their weaknesses. The scale product technique was pioneered by Eysenck

and Crown (1949) and modified by Hulka, Zyzanski, Cassel, and Thompson (1970) and by Zyzanski, Hulka, and Cassel (1974). In essence, it combines the Thurstone scale's position anchoring of an item with the Likert scale's integer item response scores. It has been shown to produce reliabilities superior to those that either the Thurstone method or the Likert method can produce separately (Castle, 1953; McNemar, 1946; Zyzanski et al.).

The scale product method entails the creation of a pool of items that are rated by judges, as described above for the Thurstone method. This judgement results in median ratings that are used as weights for each item. When the scale is administered, the subject responds to items using a balanced Likert format. The integer values assigned to the individual responses are then multiplied by the corresponding Thurstone weight to create a "weighted item score" that is summed to derive the total score. Therefore, a respondent's score for any one item is based on two pieces of information — the extent to which s/he agrees or disagrees with the statement, and the rated importance of that statement in relation to the concept being measured. Total scores can be summed for the entire instrument or for factors or other component scores, depending on the dimensionality of the instrument. The ability to use the weights within factors or other dimensions of an instrument avoids the unidimensional limitation of the Thurstone method.

The scale product method is not the only means of weighting an item. Other weighting methods include item response modelling (Hambleton, Swaminathan, & Rogers, 1991), whereby items in an affective measure are weighted according to their ability (least difficult to most difficult) to be agreed with at evaluation (Beck & Gable, 2001), and an explicit approach such as giving more weight to items that correlate with an external criterion (Rudner, 2001). Both of these approaches have advantages (e.g., item response theory allows for estimation of the "difficulty" [here, "agree attractiveness"] of items so that a developer can strive to have items that cover the range of "difficulties"), but their basis for weighting is different from that of the scale product method. The scale product method's basis for weighting — the importance of the items, as explicitly stated by the respondents, in reference to the concept being assessed — has both conceptual and cognitive appeal in the assessment of patient satisfaction. This is not to suggest that this method is superior to the aforementioned approaches to weighting; it is simply different.

The disadvantages of the scale product method stem primarily from its foundation in classical measurement (test) theory. In this theory, all items in a scale are assumed to be of equal value and are therefore summed without regard to differences in intensity or importance. Weighting of these items does not necessarily alter any statistical opera-

tions, leading some measurement experts to suggest it is not worth the effort (Gulliksen, 1950). However, knowing that the items are not of equal importance/intensity may well be enough to make the effort worthwhile.

Use of the Scale Product Technique

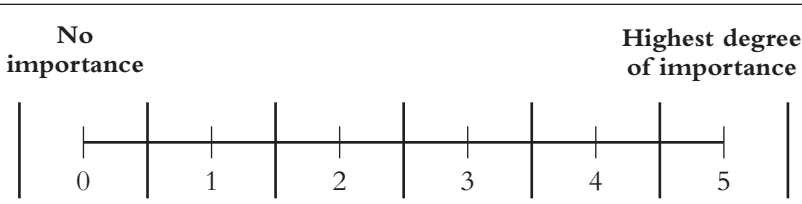
In the current study, the scale product technique was used to derive item weights for the 90 aspects of quality nursing care derived from qualitative interviews with 29 patients at two medical centres in the southwestern United States. All patients were asked the same grand tour question: "How would you describe or define good nursing care?" Further questioning depended on the response to the previous question(s) and continued until there were no unexplored areas. At the conclusion of the 29 interviews, data saturation had been achieved (Lynn & Sidani, 1991, 1995). The interviews were transcribed verbatim and the analysis began with the first interview. Ninety distinct data bits, representing attitudes ("likes being a nurse"),[1] actions ("responds to my calls promptly"), characteristics ("looks professional"), and physical environment ("my room is not noisy"), were identified in the interview data. These 90 data bits were translated into items (e.g., "looks professional" became "the nurse has a professional appearance") and presented to a panel of six patients not previously interviewed, to determine the extent to which the items matched their experience. The list was deemed accurate and comprehensive by the panel, thereby supporting the content validity of the items (Lynn, 1986). From the patients' perspective, the conceptual categories of quality nursing care were "responsiveness," "attitude," "knowing me as a person," and "respecting me as a person."

After approval had been received from the Institutional Review Board, the scale product technique was initiated by employing the Thurstone judgement stage with patients in seven hospitals in the southeastern United States — an academic medical centre, a Veterans' Administration Medical Center, three urban community hospitals, and two rural community hospitals. These hospitals were participating in a study on the perceptions of quality of care held by patients and nurses and the influences on those perceptions. The 90 aspects identified by patients as important in good nursing care were presented to 448 patients in these hospitals. The patients ranged in age from 18 to 90 years (*mean* = 53, *SD* = 16.4), were almost evenly divided between females (48%) and males (52%), and were primarily Caucasian (64%), with 65% having no more than a high-school education. Most (77%) had been hospital-

[1] Each phrase in parentheses is one of the 90 distinct data bits obtained from the interviews.

ized at least twice previously and had been hospitalized for at least 48 hours before participating in the study. All patients spoke English but were not required to read English; the data collector could read the card aloud so that the patient could then place the card on the numbered board according to her/his evaluation of the intensity of the item. See Figure 1 for the board used in this procedure.

Figure 1 *Item Sorting Board*



The participants were told that information written on each card had come from interviews with patients who were asked to describe or define good nursing care. The participants were then asked to rate the aspect described on each card in terms of its importance to the overall concept of good nursing care. The participants were given as much time as necessary to perform the card sorting and were allowed to re-evaluate any card (or stack of cards) if they chose. They took between 15 and 25 minutes to perform the card sorting.

The Thurstone method was modified in the current study by eliminating the unfavourable end of the judging continuum, since the items were all generated from interviews relating to good care. Therefore a six-point scale was used in the judgement, with panellists placing each item in one of six categories ranging from 0 (no importance) to 5 (highest degree of importance) according to their perception of quality nursing care (Figure 1). Sorting was continued until stable median rankings were achieved.

Median ratings for the 90 items ranged from 2.88 to 4.81. Aspects of care pertaining to the competence of the nurses and the adequacy of resources ranked the highest and aspects pertaining to the patient's physical environment ranked the lowest. Applying the Thurstone technique to weight these aspects of good nursing care will add to the variability and, it is hoped, the precision of the scale. Items will contribute to the total score of the scale or subscale(s) in direct proportion to their overall relevance to quality nursing care. See Table 1 for a sample of the items and their associated median ratings.

Table 1 ***Sample PPQS-ACV Items with Median Rating***

I wish the view from my room was more interesting	2.88
The nurse uses touch to reassure or support me	3.88
The nurse knows who I am as a person	3.97
The nurse makes sure that I have plenty of time to talk to her/him	4.12
The nurse shows me that I am her/his first concern	4.34
The nurse helps me take care of my daily physical needs	4.42
The nurse frequently checks on me	4.47
The nurse is patient	4.52
The nurse is able to talk to me	4.55
The nurses see me as an individual, a real person	4.56
The nurse is attentive and responsive to my needs	4.59
My room is kept clean	4.61
The nurse is clear when teaching me about my care	4.66
The nurse gives me my medications on time	4.74
The nurse knows what she/he is doing	4.81

The scale in question has five factors — Communication, Professionalism, Individualization, Timeliness, and Environment. Weighted and unweighted means and standard deviations for the factors are shown in Table 2. As might be expected, the means and standard deviations of the factors increased in the weighting procedure, by approximately the same magnitude. However, it should be noted that, as is true with any linear transformation of scores, the weighting of items has little or no effect on statistical operations, specifically those based on covariances/correlations. Only in some instances have the resulting correlations between weighted factors and outcome of interest been different from those of the original, unweighted, factor scores (Jansen, Stigglebout, Nooij, & Kievit, 2000). A few correlations in this examination were different, but by only .01 or .02. Therefore, correlations with other variables will be almost identical using weighted or unweighted factor scores, and reliability of the factors should be similar across the two scoring methods. While there were differences in the reliability estimates for the factors using the weighted and unweighted items, these differences occurred at the 3rd and 4th decimal points and therefore also are of no consequence. There are some differences in the skewness and kurtosis of the distributions of the factor scores caused by the wider dispersion of the scores created by the weighting. Despite the lack of statistical differences with the scale product technique, its primary advantage is its role in enhancing the variability of the scores.

Table 2 *Means, Standard Deviations, Skewness, and Kurtosis for Unweighted and Weighted Scale Factors*

Scale Factor (Alpha)	Unweighted				Weighted			
	Mean	*SD*	*Skewness*	*Kurtosis*	*Mean*	*SD*	*Skewness*	*Kurtosis*
Communication (.93)	4.17	.51	-.451	1.777	18.99	2.31	-.444	1.765
Professionalism (.79)	4.27	.51	-.641	2.192	19.46	2.31	-.645	2.214
Individualization (.84)	4.00	.55	-.465	1.256	16.93	2.34	-.469	1.260
Timeliness (.84)	3.99	.63	-.916	2.243	17.95	2.80	-.917	2.234
Environment (.68)	3.62	.69	-.497	.480	13.73	2.54	-.557	.620

Discussion

In essence, the scale product technique allows for subjects to consider the importance of each item when responding to it, so that the total score indicates the extent to which they agree or disagree with the item as well as the importance of the item with respect to the concept being measured. Since the score on an instrument using the scale product technique is directly proportional to the mathematical combination of the item weight and the respondent's selection, the overall importance of selected aspects is integral to the final score without the need for importance to be assessed each time the scale is used. Such an approach is particularly suited to the measurement of patient satisfaction, because it combines the evaluation of the extent to which the aspect of patient care occurred and the importance of that aspect to the patient's overall experience. Certainly there are instruments that incorporate an "importance" response scale and a "presence" or "agreement" response scale. When completing such scales, respondents are asked to provide two responses to each statement, one indicating the importance they give to the item and the other the extent of their agreement or disagreement with the statement. There are two problems with this response format: it places a large response burden on the respondent, especially when the scale is fairly long, and the importance ratings are not generalizable. Importance in such instruments is idiosyncratic.

With the genuine Thurstone scale, the items are expected to fall across the entire judgement range. In the case of the scale product technique, they are not expected to be distributed in this manner, since they already represent "important" aspects of quality nursing care. However, the fact that they varied at all suggests that even items that appear to be similar in their relative strength towards the measurement objective of an instrument may not be similar at all. It may not be reasonable, or even desirable, to assume that items in a Likert scale have the same strength, intensity, or degree of favourability towards the measurement objective. While it is conventional to treat Likert items as if they were all of equal intensity, perhaps the time has come to examine their intensity as part of the instrument's development or revision. The advantages of being able to weight items according to their strength or intensity will likely increase the sensitivity of the instrument, thereby enhancing both its validity and its reliability.

Use of a qualitative method to derive the content for the Thurstone judging method addresses a significant problem with traditional satisfaction surveys — the lack of clarity on what patients value about the care they receive (Ware et al., 1977). Weighting also helps to resolve the persistent problem of score invariance in traditional satisfaction assessments

(McDaniels & Nash, 1990; Ware, 1977). When an item is weighted according to its judged priority, the total score reflects the patient's evaluation of that aspect of nursing care, weighted for the importance of that component of care to patients in general. A respondent's agreement with an item contributes more if it is a highly weighted rather than a lesser-weighted item. Conversely, if a respondent disagrees with a highly weighted item, more points are subtracted from the score than would be the case with a lesser-weighted item. These subtle differences in the effect of items on the total score make some scales more sensitive to patient evaluation of the quality of nursing care and provide considerably more variance than traditional patient quality or satisfaction inventories.

The scale product technique has three additional advantages in terms of satisfaction scales. First, it provides a statistical check on the quality of the items. If an item receives very low rankings from the judges, it probably does not belong on the scale, as the judges have said it has little importance with respect to the concept being measured. Therefore, the weights can be used to help eliminate "bad" items. Second, giving patients the opportunity to comment on the importance of scale items can enhance the validity of the scale. Since a scale is intended to capture the perspective of patients with respect to quality of care, if patients think an item is irrelevant, then it *is* irrelevant, and is not a valid measure of quality of care from the perspective of patients. Finally, the determination of the relative importance of the items used to assess patient satisfaction offers clinicians and researchers insight into patients' relative valuing of care that otherwise is not available to them without overt investigation in this area. Therefore the scale product technique, when used with qualitatively derived items, should be useful in clinical, research, and quality improvement projects that require a patient-centred method of evaluating care.

References

Beck, C. T., & Gable, R. K. (2001). Item response theory in affective instrument development: An illustration. *Journal of Nursing Measurement, 9*(1), 5–22.

Bond, S., & Thomas, L. H. (1991). Issues in measuring outcomes of nursing. *Journal of Advanced Nursing, 16,* 1492–1502.

Castle, P. F. C. (1953). A note on the scale-product technique of attitude scale construction. *Occupational Psychology, 27,* 104–108.

Committee on Quality Health Care in America. (2001). *Crossing the quality chasm: A new health system for the 21st century.* Institute of Medicine. Washington: National Academy Press.

Corriher, J. B. (1994). *Critical care patients' satisfaction with nursing care.* Unpublished master's thesis, University of North Carolina School of Nursing.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Holt, Rinehart & Winston.

Donabedian, A. (1980). *Explorations in quality assessment and monitoring. Vol 1: The definition of quality and approaches to its assessment.* Ann Arbor, MI: Health Administration Press.

Edwards, A. L. (1957). *Techniques of attitude scale construction.* New York: Appleton-Century-Crofts.

Ellencweig, A. Y. (1992). *Analyzing health systems: A modular approach.* New York: Oxford University Press.

Elling, R. H. (1974). Case studies of contrasting approaches to organizing for health: An introduction to a framework. *Social Science and Medicine, 8,* 263–270.

Elling, R. H. (1980). *Cross-national study of health systems — political economics and health care.* New Brunswick, NJ: Transaction Books.

Eysenck, H. J., & Crown, S. (1949). An experimental study in opinion-attitude methodology. *International Journal of Opinion and Attitude Research, 3,* 47–86.

Eysenck, H. J., Crown, S., & Shapiro, M. (1950). Symposium on primary social attitudes. In *Proceedings and papers of the Twelfth International Congress of Psychology.* Edinburgh: Oliver & Boyd.

Ferguson, L. W. (1939). The requirements of an adequate attitude scale. *Psychological Bulletin, 36,* 664–673.

Fosbinder, D. (1994). Patient perceptions of nursing care: An emerging theory of interpersonal competence. *Journal of Advanced Nursing, 20,* 1085–1093.

Gulliksen, H. O. (1950). *Theory of mental tests.* New York: John Wiley.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Hancock, K., Chang, E., Chenoweth, L., Clarke, M., Carroll, A., & Jeon, Y. (2003). Nursing needs of acutely ill older people. *Journal of Advanced Nursing, 44,* 507–516.

Hilton, B. A., Budgen, C., Molzahn, A. E., & Attridge, C. B. (2001). Developing and testing instruments to measure client outcomes at the Comox Valley Nursing Center. *Public Health Nursing, 18,* 327–339.

Hulka, B. S., Zyzanski, S. J., Cassel, J. C., & Thompson, S. J. (1970). Scale for the measurement of attitudes toward physicians and primary care. *Medical Care, 8,* 429–436.

Jansen, S. J., Stiggelbout, A. M., Nooij, M. A., & Kievit, J. (2000). The effect of individually assessed preference weights on the relationship between holistic utilities and nonpreference-based assessment. *Quality of Life Research, 9,* 541–547.

Jelinek, R. C. (1967). A structural model for the patient care operation. *Health Services Research, 2,* 226–242.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 22,* 5–55.

Louden, T. L. (1989). Customer perception counts in quality assurance. *Hospitals, 63*(2), 84.

Lynn, M. R. (1986). Quantification and determination of content validity. *Nursing Research, 34,* 254–256.

Lynn, M. R., & Sidani, S. (1991, May 2–4). *The next step: The patient's perception of quality scale.* Paper presented at the 24th WSRN Annual Communicating Nursing Research Conference, Albuquerque, New Mexico.

Lynn, M. R., & Sidani, S. (1995). Patients' and family members' perceptions of what makes quality patient care. In S. G. Funk, E. M. Tornquist, M. T. Champagne, & R. A. Wiese (Eds.), *Key aspects of caring for the acutely ill: Technological aspects of patient education and quality of life* (pp. 292–298). New York: Springer.

McDaniels, C., & Nash, J. (1990). Compendium of instruments measuring patient satisfaction with nursing care. *Quality Review Bulletin, 16,* 182–188.

McIver, J. P., & Carmines, E. G. (1981). *Unidimensional scaling.* Beverly Hills, CA: Sage.

McNemar, Q. (1946). Opinion-attitude methodology. *Psychological Bulletin, 43,* 289–374.

Murdaugh, C. (1992). Quality of life, functional status, patient satisfaction. In *Patient outcomes research: Examining the effectiveness of nursing practice. Proceedings of the State of the Science Conference* (pp. 91–96). Bethesda, MD: Department of Health and Human Services, Public Health Service, National Institutes of Health.

Murphy, G., & Likert, R. (1938). *Public opinion and the individual.* New York: Harper.

Oberst, M. T. (1984). Patients' perceptions of care: Measurement of quality and satisfaction. *Cancer, 53,* 2366–2373.

Rubin, H. R. (1990). Patient judgments of hospital quality, patient evaluations of hospital care: A review of the literature. *Medical Care, 28,* S3–S9.

Rudner, L. M. (2001). Informed test component weighting. *Educational measurement: Issues and practice, 20*(1), 16–19.

Stevens, S. (1946). On the theory of scales of measurement. *Science, 103,* 677–680.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology, 23,* 529–554.

Thurstone, L. L., & Chave, E. J. (1929). *The measurement of attitude.* Chicago: University of Chicago Press.

Van Alphen, A., Halfens, R., Hasman, A., & Imbos, T. (1994). Likert or Rasch? Nothing is more applicable than good theory. *Journal of Advanced Nursing, 20,* 196–201.

Walker, J. D. (1993). Enhancing physical comfort. In M. Gerteis, S. Edgman-Levitan, J. Daley, & T. L. Delbanco (Eds.), *Through the patient's eyes* (pp. 119–153). San Francisco: Jossey-Bass.

Ware, J. E. (1977). *Effects of acquiescent response set on patient satisfaction ratings.* Report #P-5675. Washington: Department of Health, Education and Welfare.

Ware, J. E., Davies-Avery, A., & Stewart, A. L. (1977). *The measurement and meaning of patient satisfaction: A review of the literature.* Washington: Department of Health, Education and Welfare.

Yen, M., Chen, C., & Chou, S. (2003). Concerns of hospitalized care from patients' perspectives. *Journal of Nursing Research, 10*(2), 121–128.

Zyzanski, S. J., Hulka, B. S., & Cassel, J. C. (1974). Scale for the measurement of "satisfaction" with medical care: Modifications in content, format and scaling. *Medical Care, 12,* 611–620.

Authors' Note

This study was funded by the National Institute for Nursing Research, R01 NR03606.

Comments or inquiries may be directed to Mary R. Lynn, School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7460 USA. Telephone: 919-966-5450. Fax: 919-966-2286. E-mail: Mary_Lynn@unc.edu

---

Mary R. Lynn, PhD, is Associate Professor, School of Nursing, University of North Carolina at Chapel Hill, USA. Bradley J. McMillen, PhD, is Senior Research and Evaluation Consultant, North Carolina Department of Public Instruction, Raleigh.

CJNR 2004, Vol. 36 N° 3, 82–98

Résumé

Y a-t-il quelque chose qui nous échappe? La poursuite de recherches sur l'attrition

Lenora Marcellus

L'attrition, ou la perte de participants au cours d'une étude, peut mettre en péril, de façon significative, l'intégrité d'une étude longitudinale et l'élaboration de théories issues de la recherche. Bien que l'attrition fasse récemment l'objet d'un intérêt renouvelé, ce phénomène est peu signalé et peu étudié, malgré le fait qu'il peut potentiellement gauchir les résultats d'une étude. La validité interne et externe, la fiabilité et la validité statistique subissent tous les effets de l'ampleur restreinte et, fort probablement, de la nature non aléatoire d'un échantillonnage. L'élaboration d'une théorie portant sur l'attrition permettra aux chercheurs d'élaborer des stratégies d'échantillonnage qui amélioreront la qualité des données obtenues dans les études longitudinales. L'auteure propose et décrit un modèle théorique écologique relativement à la participation dans le cadre de recherches.

CJNR 2004, Vol. 36 N° 3, 82–98

Are We Missing Anything? Pursuing Research on Attrition

Lenora Marcellus

Attrition, or loss of participants over the course of a study, presents a significant threat to the integrity of a longitudinal research study and theory development resulting from the study. Although there has been a recent resurgence of interest in attrition, it is an underreported and understudied phenomenon despite its potential to introduce bias. Internal and external validity, reliability, and statistical validity are all impacted by a small sample and, most likely, a non-randomness in the study sample. Development of a theory of attrition will assist researchers in development of sampling strategies that will enhance the quality of their data in longitudinal designs. An ecological theoretical model of research participation is proposed and described.

Keywords: research participants, patient selection, research-participant relations, sample size

Nurses are increasingly being encouraged to consider longitudinal research designs as an effective way to study human experience over time (Russell & Gregory, 2000). Longitudinal research, whether quantitative or qualitative, provides an opportunity to study the nature and results of change as humans move through developmental and transitional experiences. The study of patterns and change over time is of particular relevance for nurses because of the diversity of nursing experience; this includes following the development of infants and children within families, exploring the trajectories of chronic illnesses, and identifying healthy mechanisms of ageing (Gottlieb & Feeley, 1999).

There are many advantages to choosing a longitudinal research design. These include the ability to describe patterns over time, establish the direction and magnitude of causal relationships, and describe and analyze dynamic change processes. Longitudinal designs are useful for a range of applications, including intervention, prevention, developmental, and evaluation studies. Longitudinal designs, by their very nature, also present significant challenges. They are complex and require significant commitment in terms of time, resources, and data management. Many researchers agree that participant attrition is one of the greatest disadvantages of longitudinal research. Even with the most sophisticated designs, participant attrition is considered to be a hindrance to research fidelity (Gross & Fogg, 2001; Mason, 1999).

We may stand to learn just as much from those who do not complete a study as from those who do. In this article I review the latest research on attrition, from the perspective of maintaining study integrity and consequently developing knowledge. I argue the position that attrition needs to receive more attention in research studies from theoretical, design, analytical, and reporting perspectives. I propose a theoretical model of attrition as a way of organizing considerations of barriers to research participation and strategies to maximize participation. Throughout, I will provide examples of longitudinal studies related to health, in particular intervention studies, to illustrate the points of discussion.

Definition, Prevalence, and Effects of Attrition in Longitudinal Studies

Within the literature, there are inconsistent definitions of attrition, resulting in a range of reported variability in attrition rates. Broadly, attrition can be defined as the failure of subjects to complete their participation in a study following enrolment (Given, Keilman, Collins, & Given, 1990). Participants may be considered non-completers if they do not complete the treatment protocol or if they miss any data-collection point during the study. Attrition is generally seen towards the end of a study, but sometimes it is seen throughout the study, with cumulative loss occurring with successive data-collection sweeps.

Reported attrition rates range from 5 to 70%. Although there is no absolute standard for acceptable attrition levels, bias is generally a concern if the rate exceeds 20% (Polit & Hungler, 1995). Boyle, Offord, Racine, and Catlin (1991) reviewed studies on childhood psychiatric disorders and found a general loss of 20 to 30% of participants from the first follow-up. In studies of adolescent substance use, losses of 20 to 55% are common. Furthermore, many studies do not even report on attrition. Goodman and Blum (1996) studied the organizational literature and found attrition rates ranging from 0 to 88%, with a median of 27% — and 44% of the studies did not mention attrition at all. Sifers, Puddy, Warren, and Roberts (2002) examined 260 research articles published in child psychology and child development journals and found that attrition rates were significantly underreported. The recent nursing research literature includes no similar studies. However, many nursing research journals follow the guidelines of the American Psychological Association, and since the fifth (and latest) edition of the APA's *Publication Manual* recommends the reporting of attrition information, we can expect to see increased reporting of attrition.

Reporting requirements for researchers have been developed and continue to be refined, particularly for those involved in large clinical

trials. For example, the CONSORT (Consolidated Standards of Reporting Trials) consensus statement on recommendations for improving reporting on the quality of randomized trials, developed by an international group of clinical trialists, statisticians, epidemiologists, and biomedical editors, includes attrition (called participant flow) as an item to report (Moher, Schultz, & Altman, 2001).

With the increasing complexity and costs of coordinating and conducting large longitudinal research studies, it is timely that participant attrition is the subject of renewed attention. Attrition has the potential to introduce significant bias into a study and represents a threat to internal, external, and statistical validity. Differential attrition may lead to selection bias, which is a threat to internal validity in comparative studies. External validity is also affected by sample bias, as completers may differ from noncompleters on various characteristics. Consequently, the sample may not be truly representative and extrapolation beyond the study may be impossible. Significant attrition may also result in a reduced sample size, leading to low statistical power.

Typical Management of Attrition

Farrington (1991) finds it ironic that more books and articles have been written about technical problems such as research design and statistical analysis than about practical problems such as attrition that may be more difficult to resolve and more consequential for the validity of conclusions. Researchers may see attrition as a practical rather than theoretical problem, and therefore accord it less prestige (Murphy, 1990).

Rather than anticipating and minimizing attrition, researchers have historically accepted it as part of the normal course of a longitudinal study and then focused on detecting and compensating for it. The method chosen to deal with attrition usually depends on the pattern of data that are missing (Kneipp & McIntosh, 2001). Additional statistical analyses are then incorporated into the study design to compensate for missing data and to increase confidence in the results.

The estimation of missing data has become a legitimate research subfield in and of itself. The increasing sophistication of statistical computer programs has advanced the management of missing data. Briefly, a few decades ago there were few methods available to manage missing data. The classic approach was to "get rid of" the data by doing listwise or pairwise deletions, or by just not using the data set (Patrician, 2002). Newer methods include single and multiple approaches to imputation. The most frequently used method in longitudinal studies is multiple imputation, which allows the researcher to use the existing data to impute values generated from the analysis of the data set. Multiple impu-

tation approximates the "real" value while preserving the uncertainty of the missing values (Patrician).

Statistical procedures for missing data continue to improve. Schafer and Graham (2002) suggest that newer methods such as multiple imputation, latent variable programs, and Full Information Maximum Likelihood are preferable to some of the older procedures. Most approaches to the handling of missing data are based on the assumption that data are missing at random. For data that are not missing at random, there is currently no reliable method of correction. However, Schafer and Graham report on upcoming statistical advances that will make it possible for researchers to deal with missing data that are not missing at random. Harris (1998) identifies a need for research to continue trying to develop appropriate analyses for non-random data; at present, a "study plagued by selective attrition cannot be rescued" (Hirdes & Brown, 1994, p. 349). Finally, an alternative to simply dealing with attrition is to plan for it. Graham, Taylor, and Cumsille (2001) propose that researchers consider building planned "missingness" directly into longitudinal studies. Researchers must still make every effort to maintain high participation rates and to minimize the effects of selective attrition.

The focus of the attrition research literature lies within the context of quantitative studies. Longitudinal qualitative research is considered a developing methodology and, other than retrospective biographical or life history methods, there are a limited number of published longitudinal qualitative research studies (Thomson & Holland, 2003). Despite the fact that this is an emerging method, attrition has already been identified as a key threat to the integrity of a study. Thomson, Plumridge, and Holland (2003) identify participant retention as the "most obvious imperative" of this method. Accordingly, they express concern that the method is developing without a relevant literature to inform and debate the epistemological and practical decisions made during development and conduct of the studies. Although the matter is beyond the scope of this article, there is a need for theoretical discussion of the appropriateness and application of the concept of attrition (as understood within quantitative methodologies) to the qualitative paradigm.

Variation in Attrition

Goodman and Blum (1996) suggest that attrition is affected by three types of variables: participant, researcher, and contextual. Participant variables, which usually include demographics, are often studied and reported. Some researchers, however, have found that emotional, psychosocial, and contextual factors are among the better predictors of attrition. For instance, non-completers may feel that their health has

improved and therefore they do not need follow-up. In a study of child health, Boyle et al. (1991) found that their largest sample loss occurred among children with psychiatric disorders living in adverse family situations. In her study of an adolescent parenting program in a neonatal intensive care unit, Letourneau (2001) noted that mothers who had partner difficulties and who visited their infants less frequently were more likely to miss follow-up visits.

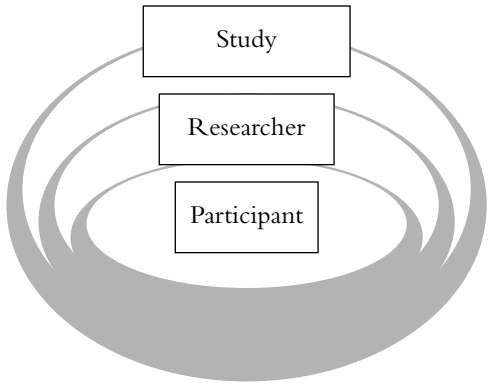
So far little direction has been provided for understanding why participants leave studies or clients leave treatment programs. Many studies of attrition are simply byproducts of studies designed to answer other research questions (Given, Given, & Coyle, 1985). Harris (1998) suggests that the study of attrition has shown an over-reliance on simple and atheoretical analyses, which highlights the absence of thoughtful, well-constructed, well-conceived theories of the underlying causes of attrition. She treats the study of attrition as a legitimate research sub-field. Flick (1988) presents an intriguing perspective on attrition. She suggests that rather than approaching it as a nuisance, researchers should treat it as a legitimate part of the phenomenon of interest and as an outcome variable.

The causes or correlates of attrition itself may be theoretically important and could be routinely built in as attributes that are measured and tested. In the quest to learn how factors, both singly and in combination, lead to attrition, a set of norms may evolve that will lead to the development of a theory of attrition that researchers can use in designing strategies to minimize attrition (Given et al., 1985). In the next section I will present a theoretical model generated from the literature on attrition. It reflects an initial step in the development of a theory of attrition.

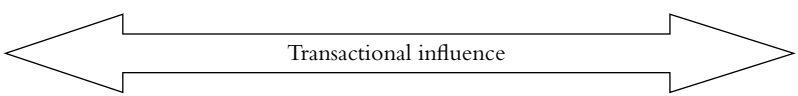
An Ecological Theory of Research Participation

Goodman and Blum's (1996) broad consideration of attrition, particularly because of the inclusion of context, invokes the notion of ecological developmental theory. Similarly, Given et al.'s (1985) call for the development of a conceptual model of attrition that addresses the interaction of multiple variables, including the interaction between intervener and participant, supports an ecological approach. Ecological perspectives stress the importance of interaction between individuals and their environment. An ecological perspective of research participation is proposed as a model within which to design and study multilevel retention strategies. This model consists of a series of nested layers that represent various influences on an individual's ability and desire to participate in a study (Figure 1). The factors within each layer are drawn from longitudinal studies of attrition. A large number of correlates of retention were iden-

Figure 1 ***An Ecological Theory of Attrition***



Participant	**Researcher**	**Study**	**Environment**
Motivation	Motivation	Complexity of design	Philosophical
Values	Personal meaning	Perspective on subject	Disciplinary
Beliefs	Values	Participation	Organizational
Personal meaning	Beliefs	Perceived importance	Practice
Specific population characteristics	Communication style		Political
	Respect for participants		Geographical
			Funding



PARTICIPANT-CENTRED

tified in these studies, including characteristics specific to the participants, the researchers, the individual studies, and the environment (philosophical, disciplinary, organizational practice, political, funding) in which the study took place (Harris, 1998).

In addition to the core layers, two key influences affect all levels of attrition. The first is transactional, reciprocal, and interactive — that is, each layer has a direct influence on the action of the system immediately adjacent to it and an indirect transactional influence, through that adjacent system, on the actions of all other systems (Drummond, 2004). The

second relates to the participant-centred approach to research, which is considered necessary. A theory of attrition that is relevant to the current research environment is one that recognizes and capitalizes on the increasing activism of people who are potential research participants (Gross & Fogg, 2001; Mitchell, 2001). Researchers are encouraged to move from a paternalistic relationship with participants to a relationship that is participant-centred (Gross & Fogg).

The ecological theory of research participation would give investigators the opportunity to address situation-specific barriers and develop strategies specific to the study and the population of interest in order to maximize participation. It would serve to guide the assessment of why participants do or do not remain in a study and what environmental or community factors might interact with the participant, researcher, or study to affect attrition.

In addition to considering recruitment and retention issues, which focus on timing and action, an ecological approach would incorporate meaning and interaction, which reflect the key influences of transaction and centring on the participant within the research relationship. Gross and Fogg (2001) suggest that the goal of such an approach is to reach the highest level of scientific inference while still acting in the best interests of the participants.

Two cautions are necessary with regard to this process of reframing attrition. First, the researcher must take care to maintain a balance between individualizing and contextualizing retention strategies and to ensure that retention strategies do not affect the dependent variable or the study outcome (Good & Schuler, 1997). Second, the potential exists for the misuse of attrition theories. For example, clinical trial recruiters may have determined, through research, which characteristics are associated with retention and completion and therefore sample for persons who possess those characteristics. Such characteristics (e.g., ability to adhere to a complex treatment protocol or a long-term assessment) may not be prevalent in vulnerable populations, which will result in the underrepresentation of these populations in the study.

Using the Ecological Theory to Design Strategies for Managing Attrition

In this section, strategies for managing attrition related to each core layer (Table 1) will be described. In addition to strategies specific to each layer, the transactional nature of the model requires examination of the interface between layers. For example, a complex treatment protocol may not be appropriate for a participant population with few life resources, particularly if delivered by researchers with time restrictions.

Table 1 *Strategies for Reducing Attrition*

Level of Ecological Model	Strategies
Participant	• Provide clear definition of sampling population — documentation on refusals with rationale, use of cohort tree for tracking • Form participant advisory group • Ensure that research-related activities are timely, convenient, and accessible • Provide material recognition that matches the characteristics and needs of the population
Researcher	• Provide favourable presentation of study — discuss its social usefulness • Establish meaningful networks with agencies that may be points of contact for potential participants • Plan for adequate preparation and supervision of data collectors — communications skills, development of study training/protocol manual, opportunity to practise data-collection prior to start of study • Develop participant bond with the study — assure participants regarding the credibility of the study; develop tools such as logo, theme, newsletter to create a project identity • Plan for continuity of participant contact with data collectors — clear expectations, pre-contact strategies, opportunities to ask questions, newsletter, prompt response to issues, sufficient time to respond to participants' needs • Collaborative effort between researchers and participants; assure participants regarding their ability to contribute to the study — remind them that they are the experts on the research topic, give feedback to validate their contributions • Express appreciation — thank you cards, recognition of birthdays, gifts or remuneration
Study	• Detail recruitment methods specific to characteristics of targeted population • Consider individualization of retention strategies within mandate of study • Develop informed-consent procedures that provide participants with realistic expectations of the study • Develop detailed tracking techniques and database • Respect the participant's time — flexibility in scheduling data collection, expressions of appreciation, provide babysitting and transportation • Ensure that measurement strategies are viable and are not onerous for participants • Ensure that hard-to-reach populations have an equal opportunity to participate in the study

Environment	• Assess environmental factors such as transportation and community perception of research • Strengthen outreach retention methods • Monitor external influences such as policy, politics, and media representation on issues related to population of interest and research question
Sources: Given et al. (1990); Good & Schuler (1997); Lauby et al. (1996); Mason (1999); Pruitt & Privette (2001); Vander Stoep (1999).	

Strategies Related to the Participant

As previously described, retention strategies related to the participant tend to focus on demographics. Factors such as meaning, values, beliefs, and motivation have been identified as impacting recruitment and retention but are often overlooked. Given et al. (1985), in their study of attrition in an intervention program for hypertension, found that participants' beliefs and perceptions and the group to which they were assigned were better predictors of non-completion than demographic factors such as age, illness status, or socio-economic status. Perceptual variables such as individuals' understanding of their illness or participation in an intervention that provided contact and social support achieved more significance within logistical regression than demographic factors. Gross, Julion, and Fogg (2001), in their study of factors related to recruitment and retention in a prevention trial, found that goodness of fit between the personal goals of the participant and the goals of the study was highly correlated with continued participation in the study.

Retention and follow-up strategies may need to be tailored to characteristics of the population under study. An example of a research study that is not sensitive to participant characteristics is one in which the subjects are paid an honorarium but are expected to travel to a university laboratory with little regard to whether they have access to a vehicle or are mobile enough to travel. This is an important consideration for nursing research with populations who are physically, emotionally, or socially vulnerable. For example, Lauby et al. (1996) examined an intervention for HIV education at two clinic sites in Philadelphia. Although the demographics were similar at the two sites, the dropout rate was higher at the clinic situated in the more transient neighbourhood. The researchers determined that the most effective reminder strategy for this population was for outreach workers to look for the clients on the street. Similarly, Barnard, Magyary, Booth, and Eyres (1987), in a study with mothers and children at high social risk, offered tangible supports such as reimbursement, transportation, and babysitting costs in recognition of the mothers' participation in the study. Given et al. (1990), in their study

with family caregivers of elders with physical impairments, restricted their attrition rate to 4.5% by adopting several strategies: preparation of data collectors, participant bonding with the study, communication with the participants, continuity of participant contact with the data collectors, respect for participants' time, collaboration between the participant and the observer, and expressions of appreciation. These strategies focus not only on the participant, but also on the interface between participant and researcher.

Strategies Related to the Researcher

The needs of researchers have historically superseded those of participants (Gross & Fogg, 2001). Research is designed by and for scientists, and participants are expected to "respond like subjects." According to Chadwick (2001), research protocols have rarely been developed with any type of community input. However, elements of participatory action research are increasingly being incorporated into longitudinal studies. Strategies in which all participants are actively engaged in the research process are grounded in notions of participation and practicality (Brown, 2001). For example, the establishment of a participant advisory group provides an opportunity for participants and researchers to jointly design a study and protocol that are relevant and appropriate for the interests of both groups of stakeholders. As an indicator of the increased emphasis being placed on participatory approaches, the recent Blueprint for Health Research and Action (Canadian Institutes of Health Research, 2003) identifies public engagement as a key cross-institute theme for future research.

The concepts of privacy and confidentiality have acted not just as protective measures but also as barriers between participant and researcher. Some research, by virtue of its design, requires a formal separation between participant and researcher (Medical Research Council of Canada, 2003). Horsfall (1995) argues that this separation is reductive and ignores the power relations between participant and researcher. Researchers, then, need to focus on incorporating strategic elements into their retention plans, elements that foster engagement with the study, throughout its entire course and with all members of the research team. Olson and Toth (1999), in a study with mothers with addiction, found that some of the procedures they used to sustain the participation of the study population, such as bonding and follow-up, had the potential to engage chemically dependent women. This experience laid the groundwork for development of the Zero to Three program, a paraprofessional advocacy program that has had measurable success in helping high-risk drug-using mothers.

Depending on the population under study, philosophical or social differences between participants and researchers can create barriers to the development of a meaningful relationship. For example, First Nations peoples have expressed concern about researchers from other cultures coming into their communities to collect data that are not meaningful or useful for the people of the community, and then using the data to further their own agendas or careers (Smith, 1999). Boys et al. (2003) describe their use of former or current drug users as interviewers in research on drug and alcohol use as an example of how social researchers employ persons who share characteristics with the study population as a way to both increase data accuracy and develop and maintain rapport with the researcher and commitment to the study.

Finally, a number of practical factors related to retention impact the relationship between researcher and participant. Sullivan (2004) categorizes these factors as either logistical (such as lack of time or resources or an unstable research team) or personal (such as level of concern for participants or level of interest in the research question). These factors may be more important for studies that are of long duration.

Strategies Related to the Study

Models used traditionally to structure study protocols have been drawn from agricultural research and do not translate well into human health and social contexts (Gross & Fogg, 2001). Many components of the research process, such as rigorous sampling, detailed measurement, and complex treatment, place an undue burden on participants. A retention strategy related to the study itself is to develop interventions and procedures that take the needs and resources of participants into account. In addition to the development of a detailed recruitment and retention protocol, Good and Schuler (1997) recommend individualization as a key strategy, particularly for clinical studies in which researchers have to deal with continuously changing patient conditions and environmental conditions. Individualization should be used cautiously and with consideration of its effect on the experimental nature of the study outcomes.

An interface between participants and the study is the determination of what constitutes a significant outcome. What is significant to the researcher may be markedly different from what is significant to the participant. For example, participants in a cancer therapy trial may feel that the study has been successful not only because the treatment has been effective for them but also because they have played a part in research that may help others. Gross and Fogg (2001) state that participants may be less interested in the fact that they achieved a measure of improvement on a testing score than in the fact that they achieved greater ability to function independently in their home.

It may be helpful to shift perspectives in study design: instead of blaming attrition on participant factors, design studies in such a way that participants are able to complete the protocols. A body of literature that may be useful in effecting this shift is the post-colonization and indigenous research literature. This work focuses on deconstruction of the privileged Western approach to knowledge development and promotes the development of methodologies that transform research into a pursuit that gives greater voice to participants (Smith, 1999). This approach is consistent with an ecological one in that it incorporates the influences of each layer and returns ownership of research knowledge to the participants.

Strategies Related to the Environment

Not all determinants of attrition are under the researcher's control, and contextual issues have not, traditionally, been routinely considered in retention planning. In the ecological model of research participation, contextual or environmental considerations include those related to philosophical, disciplinary, organizational, practice, political, geographical, and funding issues.

Geographic location plays a role in participation rates. Jacobsen (2004) examined the relationship, and its effects on attrition, between the location of the client's residence and the location of the treatment program. He describes the "treatment ecology" of a study as including neighbourhood quality, community resources, and travel burden. Cooley et al. (2003), in a US national study of quality of life among women with lung cancer, noted geographically unique recruitment and retention issues. For example, participating researchers in the southern part of the United States encountered specific challenges such as a distrust of research, competing health and family demands, and low literacy levels. Researchers reported the need to strengthen outreach retention strategies and to adapt them to the geographically different needs of the population.

Political and media-related factors also contribute to the level of interest in research participation. Tolomiczenko and Goering (2000) encountered a high level of distrust among stakeholder groups during the design and conduct of a study describing pathways into homelessness in Toronto. Recent political actions, including regional amalgamation, cutbacks to social funding, shifting of social housing responsibilities, and increased police attention, had made staff who work with the homeless population more protective of their clients. Paine, Stocks, Ramsay, Ryan, and MacLennan (2004) investigated the impact of the alarming media coverage of the Women's Health Initiative in the United States, which reported increased health risks for women receiving long-term hormone therapy (HT). In their own HT longitudinal study in Australia, Paine

et al. found that despite extensive study-related counselling many participants chose to stop their HT. Researchers need to not only monitor conditions within their own institutions and programs, but also be cognizant of external events that may play a role in participants' perceptions of the study.

Conclusion

Loss of participants is a constant in longitudinal studies. Despite the increasing availability of sophisticated statistical tools to manage missing data, there is still no substitute for a study design that ensures a high level of participation. If the phenomenon of attrition is built into the theoretical, conceptual, and methodological design, the potential for maintaining the integrity of the study will be increased. By promoting standard reporting for attrition, we will ultimately increase our confidence in the theories developed from longitudinal studies. Researchers might consider routinely reporting all attrition-relevant data (rates, description of inclusion and termination criteria, distribution of the cases that terminated early, and why) in outcome-related studies, to help the reader assess the validity of the study and to enhance the utility of future meta-analyses. Just as the issues of ethics, gender, and culture have become standards of reporting in recent years, routine reporting of attrition data will advance the rigour of nursing research (Sifers et al., 2002).

A participant-centred approach and an ecological model of research participation can potentially be a starting point for development of contemporary theories of attrition. Ultimately, "the consequences of not pursuing attrition research will be reflected in greater numbers of treatment failures, higher program costs, and continued impairment of research interpretation" (Harris, 1998, p. 9).

References

Barnard, K., Magyary, D., Booth, C., & Eyres, S. (1987). Longitudinal designs: Considerations and applications for nursing research. *Recent Advances in Nursing, 17,* 37–64.

Boyle, M., Offord, D., Racine, Y., & Catlin, G. (1991). Ontario Child Health Study follow-up: Evaluation of sample loss. *Journal of the American Academy of Child and Adolescent Psychiatry, 30*(3), 449–456.

Boys, A., Marsden, J., Stillwell, G., Hatchings, K., Griffiths, P., & Farrell. (2003). Minimizing respondent attrition in longitudinal research: Practical implications from a cohort study of adolescent drinking. *Journal of Adolescence, 26,* 363–373.

Brown, C. (2001). Action research: The method. In P. Munhall (Ed.), *Nursing research: A qualitative perspective* (pp. 503–522). Sudbury, MA: Jones & Bartlett.

Canadian Institutes of Health Research. (2003). *Investing in Canada's future: CIHR's blueprint for health research and innovation*. Retrieved July 20, 2004, from http://www.cihr-irsc.gc.ca/e/publications/20265.shtml

Chadwick, G. (2001). Clinical trials in the 21st century: A global approach. *Research in Nursing and Health, 24,* 541–542.

Cooley, M., Sarna, L., Brown, J., Williams, R., Chernecky, C., Padilla, G., et al. (2003). Challenges of recruitment and retention in multisite clinical research. *Cancer Nursing, 26*(5), 376–386.

Drummond, J. (2004). Conceiving action, tracking practice, and locating expertise for health promotion research. *Canadian Journal of Health Promotion, 36*(1), 115–120.

Farrington, D. (1991). Longitudinal research strategies: Advantages, problems, and prospects. *Journal of the American Academy of Child and Adolescent Psychiatry, 30*(3), 369–374.

Flick, S. (1988). Managing attrition in clinical research. *Clinical Psychology Review, 8,* 499–515.

Given, C., Given, B., & Coyle, B. (1985). Prediction of patient attrition from experimental behavioral interventions. *Nursing Research, 34,* 293–298.

Given, B., Keilman, L., Collins, C., & Given, C. (1990). Strategies to minimize attrition in longitudinal studies. *Nursing Research, 39*(3), 184–186.

Good, M., & Schuler, L. (1997). Subject retention in a controlled clinical trial. *Journal of Advanced Nursing, 26,* 351–355.

Goodman, J., & Blum, T. (1996). Assessing the non-random sampling effects of subject attrition in longitudinal research. *Journal of Management, 22*(4), 627–652.

Gottlieb, L., & Feeley, L. (1999). Nursing intervention studies: Issues related to change and timing in children and families. *Canadian Journal of Nursing Research, 30*(4), 193–212.

Graham, J., Taylor, B., & Cumsille, P. (2001). Planned missing-data designs in analysis of change. *New methods for the analysis of change: Decade of behavior* (pp. 335–353). Washington: American Psychological Association.

Gross, D., & Fogg, L. (2001). Clinical trials in the 21st century: The case for participant-centered research. *Research in Nursing and Health, 24,* 530–539.

Gross, D., Julion, W., & Fogg, L. (2001). What motivates participation and dropout among low-income urban families of color in a prevention intervention? *Family Relations, 50,* 246–254.

Harris, P. (1998). Attrition revisited. *American Journal of Evaluation, 19*(3), 1–19.

Hirdes, J., & Brown, K. (1994). The statistical analysis of event histories in longitudinal studies of aging. *Canadian Journal on Aging, 13*(3), 332–352.

Horsfall, J. (1995). Madness in our methods: Nursing research and epistemology. *Nursing Inquiry, 2,* 2–9.

Jacobsen, J. (2004). Place and attrition from substance abuse treatment. *Journal of Drug Issues, 34*(1), 23–50.

Kneipp, S., & McIntosh, M. (2001). Handling missing data in nursing research with multiple imputation. *Nursing Research, 50*(6), 384–389.

Lauby, J., Kotranski, L., Feighan, K., Collier, K., Semaan, S., & Halbert, J. (1996). Effects of intervention attrition and research attrition on the evaluation of an HIV prevention program. *Journal of Drug Issues, 26*(3), 663–677.

Letourneau, N. (2001). Attrition among adolescents and infants in a parenting intervention. *Child: Care, Health, and Development, 27*(2), 183–187.

Mason, M. (1999). A review of procedural and statistical methods for handling attrition and missing data in clinical research. *Measurement and Evaluation in Counseling and Development, 32,* 111–118.

Medical Research Council of Canada. (2003). *Tri-council policy statement: Ethical conduct for research involving humans.* Ottawa: Public Works and Government Services Canada.

Mitchell, P. (2001). Leading a paradigm shift in clinical trials research. *Research in Nursing and Health, 24,* 540.

Moher, D., Schultz, K., & Altman, D. (2001). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel group randomized trials. *BMC Medical Research Methodology, 1*(2). Retrieved June 20, 2004, from http://www.biomedcentral.com/1471-2288/1/2

Murphy, M. (1990). Minimizing attrition in longitudinal studies: Means or end? In D. Magnusson & L. Bergman (Eds.), *Data quality in longitudinal research*. Cambridge: Cambridge University Press.

Olson, H., & Toth, S. (1999). Samples in research on prenatal cocaine exposure: Vexing problems and practical solutions. *Journal of Drug Issues, 29*(2), 237–252.

Paine, B., Stocks, N., Ramsay, E., Ryan, P., & MacLennan, A. (2004). Use and perception of hormone therapy following media reports of the Women's Health Initiative: A survey of Australian WISDOM participants. *Climacteric, 7,* 143–152.

Patrician, M. (2002). Multiple imputation for missing data. *Research in Nursing and Health, 25*(1), 76–84.

Polit, D., & Hungler, B. (1995). *Nursing research: Principles and methods* (5th ed.). Philadelphia: Lippincott.

Pruitt, R., & Privette, A. (2001). Planning strategies for the avoidance of pitfalls in intervention research. *Journal of Advanced Nursing, 35*(4), 514–520.

Russell, C., & Gregory, D. (2000). Capturing day-to-day aspects of living with chronic illness: The need for longitudinal designs. *Canadian Journal of Nursing Research, 32*(3), 99–103.

Schafer, J., & Graham, J. (2002). Missing data: Our view of the state of the art. *Psychological Methods,* 7(2), 147–177.

Sifers, S., Puddy, R., Warren, J., & Roberts, M. (2002). Reporting of demographics, methodology, and ethical procedures in journals in pediatric and child psychology. *Journal of Pediatric Psychology, 27*(1), 19–25.

Smith, L. (1999). *Decolonizing methodologies: Research and indigenous peoples.* New York: Zed Books.

Sullivan, J. (2004). Subject recruitment and retention: Barriers to success. *Applied Clinical Trials, 13*(4), 50–54.

Thomson, R., & Holland, J. (2003). Hindsight, foresight and insight: The challenges of longitudinal qualitative research. *International Journal of Social Research Methodology, 6*(3), 233–244.

Thomson, R., Plumridge, L., & Holland, J. (2003). Longitudinal qualitative research: A developing methodology. *International Journal of Social Research Methodology, 6*(3), 185–187.

Tolomiczenko, G., & Goering, P. (2000). The process and politics of community-based research with people currently homeless. *Psychiatric Rehabilitation Journal, 24*(1), 46–51.

Vander Stoep, A. (1999). Maintaining high subject retention in follow-up studies of children with mental illness. *Journal of Child and Family Studies, 8*(3), 305–318.

Author's Note

The author would like to acknowledge financial support in the form of a University of Alberta PhD Scholarship.

Lenora Marcellus, RN, MN, is a PhD student in the Faculty of Nursing, University of Alberta, Edmonton, Canada.

Résumé

Une exploration des préférences des patients en matière de traitements dans le cadre d'une étude sur échantillon partiellement aléatoire et contrôlé

Joyal Miranda

L'étude sur échantillon partiellement aléatoire et contrôlé (ÉÉPAC) est une conception modifiée dont le but est de tenir compte des préférences des participants en matière de traitement. Les connaissances concernant les facteurs qui influencent ces préférences sont très limitées. Cette étude évalue le degré de consentement des participants à collaborer à un processus aléatoire ainsi que les facteurs qui influent sur leurs préférences en matière de traitement. Les données quantitatives et qualitatives cueillies dans le cadre d'une ÉÉPAC de deux interventions comportementales dans des cas d'insomnie ont été analysées. Parmi les participants, 75 % ont refusé d'être aléatoirement assignés à un groupe thérapeutique. Ce chiffre a grimpé à 90 % après l'introduction de renseignements concernant les deux interventions. L'information que détenaient les participants concernant le traitement, la pertinence de celui-ci et leur capacité d'y adhérer selon leur perception ont influencé leur préférence en matière de traitement. Quelques participants qui n'avaient pas de préférence marquée et qui nécessitaient un traitement étaient disposés à être assignés de façon aléatoire. Les résultats soulèvent des questions concernant l'impact de la préférence en matière de traitement et de la procédure d'assignation sur la validité des conclusions de l'étude.

Mots clés : préférence en matière de traitement, ÉÉPAC

CJNR 2004, Vol. 36 Nº 3, 100–114

An Exploration of Participants' Treatment Preferences in a Partial RCT

Joyal Miranda

The partial RCT is a modified design meant to account for participants' treatment preferences. Little is known about the factors that influence such preferences. This study investigated the extent to which participants are willing to be randomized and the factors that affect their treatment preferences. The quantitative and qualitative data collected as part of a partial RCT evaluating 2 behavioural interventions for insomnia were analyzed. Of the participants, 75% were unwilling to be randomly allocated to a treatment group. This figure increased to 90% after information on the 2 interventions was provided. Participants' knowledge of the treatment, its suitability, and their perceived ability to adhere to it influenced their treatment preference. A few participants who had no strong preference and who required treatment were willing to be randomized. The findings raise questions about the impact of treatment preference and allocation procedures on the validity of study conclusions.

Keywords: treatment preference, random assignment, partial RCT

Introduction

Demonstrating the effectiveness of interventions is a prerequisite for their use in practice. Effectiveness is concerned with the robustness of the intervention outcome under actual conditions and with different clinical populations (Sidani, & Braden, 1998; Whittemore & Grey, 2002). Studies to evaluate the effectiveness of interventions focus on the clinical applicability of the treatment under investigation and seek to duplicate the situations that clinicians will encounter in their practices (Streiner, 2002; TenHave, Coyne, Salzer, & Katz, 2003). Of the situations that clinicians encounter, selection of the most appropriate treatment for individual patients reflects an important aspect of patient-centred care. Patient-centred care (PCC) is viewed as an approach that results in high-quality care (Attree, 2001; Larrabee & Bolden, 2001). Its essence is to view each patient as a unique person, respect patients' values and beliefs, and respond flexibly to patients' individual needs and preferences (Lauver et al., 2002; McCormack, 2003). Application of PCC implies that nurses assess patients' needs, values, and preferences; inform patients of available treatment options and the risks and benefits of each; ask patients which

treatment they prefer; and provide patients with the treatment of their choice (O'Connor, Mulley, & Wennberg, 2003; Tunis, Stryer, & Clancy, 2003). Interventions that are consistent with patients' preferences are posited as beneficial, as they increase patients' sense of personal control, patients' satisfaction with care, and the achievement of desired outcomes (Cahill, 1996).

Patient preference is the expression of the value of various courses of action following informed deliberation on their risks and benefits (Bowling & Ebraham, 2001). Inquiring about patients' treatment preferences is necessary to guide practice. Knowledge about which interventions have been perceived by patients as acceptable, suitable, and desirable assists the practitioner in selecting which ones to discuss with and recommend to patients. Interventions that have been found to be effective but viewed by patients as unacceptable, unsuitable, or undesirable will not be adhered to in everyday life (Morin, Gaulier, Barry, & Kowatch, 1992; Vincent & Lionberg, 2001), thereby hindering the achievement of desired outcomes (Bradley, 1993). It is therefore important that patient preferences be assessed and accounted for in the evaluation of an intervention's effectiveness.

The randomized control trial (RCT), often considered the gold standard for evaluating the effectiveness of interventions, does not take patient preferences into account. Random assignment of participants to the experimental or control group is the key feature of RCT design. This is the most critical safeguard of internal validity; however, it ignores the participants' desires and preferences regarding the treatment options included in the study. Several authors propose alternative research designs in which the participants' treatment preferences are examined (e.g., Lambert & Wood, 2000; McPherson, & Chalmers, 1998) and taken into account in the allocation of participants to a treatment group (e.g., Brewin & Bradley, 1989; Coward, 2002). The partial RCT is proposed as a modified RCT design in which patients' treatment preferences are taken into consideration while the rigour of the RCT design is maintained (Torgenson & Sibbald, 1998). The partial RCT allows participants to say whether they would rather be randomly allocated to a treatment option or allocated to their preferred treatment.

Little is known about the process underlying participants' preferences regarding method of allocation and treatment options offered in effectiveness studies (Horne, 1999). The purposes of the present study were to (1) determine the extent to which participants are willing to be randomly assigned to a treatment before and after being informed about the options, and (2) explore the factors that influence participants' treatment preferences. The data for the study were obtained from a large partial RCT evaluating the effectiveness of two behavioural interventions, stim-

ulus control instructions (SCI) and sleep restriction therapy (SRT), in managing insomnia.

Literature Review

Accounting for participants' treatment preferences, as is done in a partial RCT, has advantages and disadvantages. These are discussed at the methodological and clinical levels.

Several studies have investigated patients' treatment preferences. Overall, the results indicate that most participants (60–79%) express a preference for a particular intervention (e.g., Awad, Shapiro, Lund, & Feine, 2000; Morin et al., 1992; North-West Uro-Oncology Group, 2002; Torgerson, Klaber-Moffet, & Russell, 1996; Vincent & Lionberg, 2001; Walter, Vincent, Furer, Cox, & Kjernisted, 1999). Accounting for such preferences has several methodological advantages. It makes recruitment of participants and achievement of the required sample size easier, as compared to the traditional RCT. The results of some studies indicate that many people who are unwilling to be allocated to a treatment based on chance refuse to take part in an RCT (King, 2000). People may decline to participate in an RCT if they highly favour one treatment option, are aware that they have only a 50% chance of receiving it, and wish to avoid receiving the less desired treatment (Bradley, 1993). For instance, O'Reilly, Martin, and Collins (1999) found that only one of 20 patients agreed to be randomized to either of two treatments for prostate cancer. They report that patients who were well informed about the two treatments refused random assignment, preferring to make their own treatment choice. Thus, the rather small number of persons who agree to participate reflect a subgroup of the target population who are willing to take the risk of receiving any treatment. Informing potential participants that their treatment preferences will be taken into consideration is an effective strategy for increasing the rate of participation in a study and facilitates achievement of the required sample size. Furthermore, the resultant sample is representative of the target population and the results are generalizable to various subgroups of the population (TenHave et al., 2003).

In contrast to random assignment, accounting for participant preferences can minimize attrition and promote adherence to the treatment protocol. Random assignment may result in a mismatch between the treatment option that participants prefer and that to which they are randomly allocated. Participants who do not receive their preferred treatment may become demoralized, disappointed, uncooperative, noncompliant, and prone to drop out of the study. Attrition reduces the sample size, which decreases the statistical power to detect significant interven-

tion effects. Offering participants the treatment of their choice improves their satisfaction with and adherence to the treatment they receive and reduces attrition (Corrigan & Salzer, 2003; TenHave et al., 2003). Retention of a large number of participants and adherence to the intervention increase the statistical power to detect significant intervention effects (Lipsey, 1990; Shadish, Cook, & Campbell, 2001).

Accounting for patient preferences in a research study may lead to results that are clinically useful and that provide a profile of participants who have strong preferences for a particular treatment and who benefited most from the intervention under evaluation. Equipped with this type of knowledge, nurses can provide care that is based on the best available evidence while responding to their patients' values and wishes (O'Connor et al., 2003). The benefit of providing care in accordance with patient preferences is illustrated in a study conducted by Ruland (1998). This study found that when nurses were given information on their patients' preferences regarding self-care goals, 74% tailored their care to those preferences, which, in turn, resulted in the achievement of self-care goals.

In contrast, accounting for participants' treatment preferences in a partial RCT has methodological disadvantages. First, it requires a large sample. A large sample is necessary to ensure adequate statistical power to compare the four groups of participants: those assigned to the experimental group and are either satisfied or disappointed with the treatment received, and those assigned to the control group and are either satisfied or disappointed with the treatment received. The four groups may vary in their response to the intervention and the achievement of the expected outcomes (Bradley, 1993; Corrigan & Salzer, 2003; McPherson & Britton, 2001; TenHave et al., 2003). Second, if participants are allowed to choose the method of assignment to a group (random or preference) and the treatment option, then there is an increased possibility of an unbalanced design resulting — that is, an unequal number of participants in the four groups (Corrigan & Salzer). Several strategies can be used to address this limitation, such as randomly selecting a subsample from the larger group size to equal the smaller group size, conducting the analysis with the unequal and the equal group sizes, and comparing the results to determine whether the between-group differences are due to unequal group size. Third, the assignment of participants to their preferred treatment option may increase their expectation that the treatment will be effective and cause them to respond accordingly, resulting in improved outcomes (Corrigan & Salzer). The evidence supporting this limitation is inconclusive. In some studies treatment preference was associated with the achievement of outcomes (e.g., Thomas, Craft, Paterson, Dziedzic, & Hay, 2004), but in other studies it had no impact on outcomes (e.g., Klaber-Moffett et al., 1999).

There are a few published studies investigating participants' treatment preferences using a partial RCT (e.g., Coward, 2002; Thomas et al., 2004). However, the reports of these studies provide limited information on the process underlying such preferences. Specifically, the factors that shape the preferences are not clear or explicit. Yet these factors would assist in the identification of the specific information that patients need during the process of forming their treatment preferences, which is a requirement of PCC. This study represents an attempt to describe this process in terms of the factors that influence participants' preferences with regard to the allocation procedure and the treatment options.

Study Design

Data for this study were obtained from a large partial RCT evaluating the effectiveness of two behavioural interventions for managing insomnia. The partial RCT design consisted of asking participants about their preferences for either of the two interventions and assigning them accordingly. A questionnaire was used to elicit their preference with regard to the allocation procedure and the intervention under investigation. The quantitative and qualitative data collected in this questionnaire were used to address the objectives of the study.

The questionnaire was administered by the researcher, in an individual interview format, after the participants consented to take part in the study and completed the pretest measures. The interview proceeded as follows. First, the participants were asked whether they were willing to be randomly allocated to a treatment group. Second, they were informed, using a script, of the nature, effectiveness, and disadvantages of each intervention. Third, they were requested to evaluate each intervention on its acceptability, suitability, and effectiveness in managing their insomnia. Finally, they were asked which intervention they preferred and how they wished to be assigned to one of the two interventions — that is, either randomly or based on their preference. Participants with a preference were allocated to their preferred/selected intervention, while those with no preference were randomly allocated to a treatment group by means of an opaque, sealed envelope. The researcher documented, verbatim, each participant's verbal responses throughout the interview. The participants' quantitative answers to the questions eliciting their qualitative verbal responses comprised the data for understanding the process underlying their preferences.

Sample

The convenience sample consisted of 67 participants. Persons were eligible for the study if they (1) lived in a non-institutional dwelling in the

community, (2) were 21 years of age or older, (3) were able to read and write English, and (4) had a complaint of insomnia that met the criteria for disorder regarding initiation or maintenance of sleep — that is, sleep onset latency and/or time awake after sleep onset of 30 minutes or more per night for a minimum of 3 nights per week, of 3-month duration or longer, as corroborated by sleep diaries and self-report. Exclusion criteria included cognitive impairment as ascertained by a Mini-Mental State Exam score of over 27 and the presence of severe psychological impairment evidenced by the Brief Symptom Inventory Global Severity Index T-score of under 50. The behavioural intervention for managing insomnia requires the active participation of the client in making the behavioural changes, which may be hindered by cognitive or severe psychological impairment.

Variables and Measures

A treatment-preference questionnaire was developed by the investigators to elicit information about the participants' preferences regarding treatment allocation method (i.e., random or based on preference) and the two behavioural interventions for managing insomnia (i.e., SCI or SRT). The questionnaire consisted of four parts. The first part inquired whether the participant was, in general, willing to be randomly assigned to a treatment option in a research study. The second part described one intervention for managing insomnia in terms of its name, nature, effectiveness (based on available empirical evidence), and disadvantages. Following the description, the participants were asked to rate the intervention for suitability, acceptability, effectiveness, and their willingness to adhere to it. The third part included a description of the other intervention and the rating scales to evaluate it. In the fourth part, participants were asked how they would like to be assigned to one of the two treatment options described in the second and third parts — that is, based on chance or on preference. The responses to the questions in the first and fourth parts of the questionnaire formed the quantitative data analyzed in this study to determine the extent to which participants were willing to be randomly assigned to treatment. In the last part of the questionnaire, the participants commented verbally on their choice. The researcher recorded these comments in writing. These responses formed the qualitative data that were content analyzed to identify the factors that influenced preferences with regard to the intervention.

Procedure

The study protocol was approved by the Research Ethics Board at the University of Toronto. When obtaining participants' consent, the research assistant (RA) described the purpose of the study, the data-collection

procedure, participants' rights as human subjects, the two behavioural interventions offered in the study, and the procedure for treatment allocation (i.e., the participants were told that the two interventions demonstrated efficacy and that they could choose the one they wished). After obtaining consent and pretest data, the RA administered the treatment preference questionnaire. The RA read the questions and the response options, and recorded the option selected by the participant. The RA did not provide any information not available on the questionnaire and did not discuss the participant's choice. After rating each intervention, the participants indicated how they wished to be allocated. Those who chose to be randomly assigned learned about their assignment by opening a sealed envelope. Those who were not willing to be randomly assigned were asked which of the two treatment options they preferred and were then assigned to an experimental group based on their preference. The participants' verbal responses upon learning of their group assignment were recorded at the end of this data-collection session.

Data Analysis

Descriptive statistics were used to characterize each participant's profile and responses to the treatment preference questionnaire. In reviewing participants' responses to the initial question (i.e., first part of the questionnaire) and the final question (i.e., fourth part of the questionnaire) concerning treatment allocation, the participants were classified into four groups: (1) those who consistently chose not to be randomly assigned, (2) those who consistently chose to be randomly assigned, (3) those who changed their decision from random to preference allocation, and (4) those who changed their decision from preference to random allocation. Within each group, verbal responses were content analyzed. The emerging themes reflected factors that influenced the participants' preferences regarding the method of allocation and the interventions.

Results

The results are presented in relation to the participants' profiles, the participants' responses to the treatment preference questionnaire, and the factors that influenced their preferences. The majority (69%) of the 67 participants were women, with a mean age of 45 years (SD = 16). Most (42%) were married, well-educated (mean years of education = 17, SD = 4), and employed either part-time or full-time. They rated their insomnia as severe, bothersome, and interfering with daily functioning. Their perceived severity of insomnia was supported by mean scores on sleep onset latency (*mean* = 50 minutes, SD = 36) and wake after sleep onset (*mean* = 54 minutes, SD = 39).

When initially asked about their willingness to be randomly assigned to a treatment group, 50 participants (75%) responded that they were not willing to be randomly assigned. However, after learning about the two particular treatment options offered in the study and rating each, 60 participants (90%) chose not to be randomly assigned. Thus, 12 participants (18%) changed their mind about how they wanted to be assigned once the information on the interventions was provided, and 55 (82%) did not change their mind. Of those who changed their mind, some selected random allocation first and allocation based on preference second, while others selected preference first and random allocation second. Of those who did not change their mind, some consistently selected random allocation and others consistently selected allocation based on preference. The distribution of the participants across the four groups is presented in Table 1.

Table 1 ***Distribution of Participants in Method of Treatment Allocation***

Group	*N* (%)	Themes
1. Consistently chose treatment allocation based on preference	49 (73)	Suitability (*n* = 27) Acquired knowledge (*n* = 20) Familiarity (*n* = 18) Compliance (*n* = 17) Control (*n* = 16)
2. Consistently chose randomization	6 (9)	No preference (*n* = 6) Just wanted treatment (*n* = 3)
3. Changed from randomization to preference	11 (16)	Acquired knowledge (*n* = 11) Control (*n* = 6)
4. Changed from preference to randomization	1 (1)	Indiscretion (*n* = 1) No preference (*n* = 1)

The themes that emerged during content analysis of the verbal responses identified factors that influenced the choice of allocation procedure. For the first group, five themes explained the consistent choice of treatment allocation based on preference. First, these participants tended to base their choice of treatment on its perceived suitability to their particular sleep problem. One person stated: "The [first] treatment sounds better for me since it deals with a number of…strategies. The [second] treatment looks at time scheduling. I'm very consistent with my bedtimes. I don't think that's my problem." Second, they rejected random

assignment because they preferred to rely on their acquired knowledge of the treatment options rather than on chance. One participant stated: "With the information I got, and knowing myself and the sleep problem I have, I would rather choose myself than leave it to a 50/50 chance; that makes no sense to me." Third, some participants based their decision against random assignment on their familiarity with one of the treatment options — that is, they did not want to risk being allocated to the treatment they were familiar with and had previously received; they wanted to receive the other, new, treatment. Fourth, these participants rejected random assignment based on their perceived ability to comply with one treatment better than the other — that is, participants who believed they were unable to adhere to a treatment protocol wanted to make sure they were not allocated to that treatment. This theme is captured in the following comment: "My life is a busy one. I'm constantly on the go. My sleep schedule often changes based on my plans for the day, so I know the [second] treatment is not for me. I wouldn't be able to follow it." Finally, some participants seized the opportunity to have control over the decision-making process and selected allocation based on preference.

For the second group — those who consistently chose to be randomly assigned — two themes emerged. One theme reflected lack of a strong preference for a treatment option. Within this group, the majority ($n = 6$) did not have a preference for either of the treatment options. After receiving information on the two interventions, they felt they had "not tried" or were "not familiar" with either and therefore did not care which one they received. The second theme related to the need for treatment. Some participants ($n = 3$) did not care about the method of allocation. They just wanted some type of treatment for their sleep problem. One participant explained: "I need a good night's sleep. I'm desperate. I'll do either one. I just want some help."

For the third group — those who changed from random to preference allocation — two themes emerged. These were similar to those identified for the first group, the participants who consistently chose not to be randomly assigned. First, the members of this group changed their mind once they received information on the two treatment options. Soon after learning about the nature, effectiveness, and disadvantages of each intervention, they developed a preference for one over the other. This preference was related to the person's familiarity with one treatment. They felt they had already learned everything there was to know about it. Also, these participants developed a preference for one treatment option based on its suitability to their particular sleep problem. One participant said: "I know what I should and should not be doing to help my sleep, but it's not working. So I'll have to go with [this intervention] and see if [it] helps." Second, the issue of control arose in this group. The par-

ticipants explained that reliance on chance to place them in a treatment option did not consider what was best for them. They preferred to choose a treatment, based on their knowledge of their sleep problem and the information obtained about the treatments. One participant said: "I'd rather rely on myself to make the decision with the information I have…than [take] a 50% chance."

The last group consisted of one participant who changed the response to the method for treatment allocation from preference to random assignment. This person showed indecisiveness and lacked a strong treatment preference. After obtaining the information about the two interventions, the participant was still indecisive as to which one would be more beneficial and therefore chose to be randomly assigned. This participant felt no real connection to either of the two treatments.

Discussion

The results indicate that the majority of participants (75%) would not like to be randomly assigned to a treatment option in any research study. This finding is consistent with that reported by O'Reilly and colleagues (1999) and supports the observation that participants are not willing to be randomly assigned to treatment groups (King, 2000). The proportion of participants who refused random allocation increased to 90% after they were informed of the nature, effectiveness, and disadvantages of the two options. Knowledge of the treatment option was one factor influencing the participants' preference with regard to the options and non-random allocation to treatment. This finding replicates the observation by O'Reilly and colleagues that well-informed patients increasingly refuse randomization and choose to make their own treatment decision. The factors that influenced the participants' choice of treatment included suitability, compliance, and familiarity with the intervention. This finding is supported by the conclusion of Morin et al. (1992) and Vincent and Lionberg (2001), who suggest that patients will not choose or adhere to interventions that are unacceptable, unsuitable, or undesirable. The issue of control in selecting a treatment has been cited as a factor in a person's decision to participate in an RCT and in treatment. Those who have a clear idea about the treatment options under investigation are less willing to leave group assignment to chance (Bradley, 1993; Ellis, 2000; Jenkins & Fallowfield, 2000).

In summary, the results of this study provide preliminary evidence indicating that participants in intervention evaluation studies do not wish to be randomly allocated to a treatment group. These findings elucidate the factors that influence a participant's decision regarding the allocation

procedure. Knowledge of the treatment options apparently informs participants of not only the nature of the treatment but, more importantly, its suitability to their problem and the extent to which they will be able to adhere to it. These factors appear to play a key role in determining their selection of allocation procedure.

These results are based on a rather small sample of persons seeking non-pharmacological treatment for their insomnia. In order to enhance generalizability, they should be replicated with a larger number of participants with different clinical conditions and seeking various types of interventions. Nonetheless, they provide some empirical support for the clinical observations that patients do have treatment preferences and that these preferences are shaped by the perceived suitability of each treatment to their condition and the extent to which their lifestyle permits adherence to it. Nurses can discuss the suitability of interventions and the patients' ability to adhere to treatment when eliciting their preferences, which is an important aspect of PCC. Further, the results show that participants who have no strong preference for a treatment option and participants who perceive a need for treatment are willing to be randomly assigned. It can therefore be concluded that the factors that influence the decision to be allocated based on preference differ from those that influence the decision to be allocated based on chance. The extent to which these factors directly or indirectly affect adherence to treatment, and whether any placebo effect associated with the expectation of improved outcome resulting from the chosen treatment, is not known and should be investigated in future studies.

Conclusion

When given the option of not being randomly allocated to a treatment group and presented with information on the treatments being offered, few participants in intervention studies are willing to be randomly allocated; most tend to choose a treatment based on preference. Information on the treatment and perceptions about its suitability and one's ability to adhere to it, as well as the desire to have a role in decision-making, are factors that affect participants' treatment preferences and their decision with regard to allocation. Therefore, it may be important for researchers conducting intervention studies to account for participants' treatment preferences and to determine the extent to which these preferences influence the outcome of the treatment and the patient's satisfaction with it. Examination of the extent to which patient preferences affect outcomes enhances the validity and clinical relevance of findings in effectiveness research.

References

Attree, M. (2001). Patients' and relatives' experiences and perceptions of "good" and "not so good" quality care. *Journal of Advanced Nursing, 33*(4), 456–466.

Awad, M. A., Shapiro, S. H., Lund, J. P., & Feine, J. S. (2000). Determinants of patients' treatment preferences in a clinical trial. *Community Dental Oral Epidemiology, 28*(2), 119–125.

Bowling, A., & Ebraham, S. (2001). Measuring patients' preferences for treatment and perceptions of risk. *Quality in Health Care, 10*(Suppl 1), 12–18.

Bradley, C. (1993). Designing medical and educational intervention studies. *Diabetes Care, 16,* 509–518.

Brewin, C. R., & Bradley, C. (1989). Patient preferences and randomized clinical trials. *British Medical Journal, 299,* 313–315.

Cahill, J. (1996). Patient participation: A concept analysis. *Journal of Advanced Nursing, 24,* 561–571.

Corrigan, P. W., & Salzer, M. S. (2003). The conflict between random assignment and treatment preference: Implications for internal validity. *Evaluation and Program Planning, 26,* 109–121.

Coward, D. (2002). Partial randomization design in a support group intervention study. *Western Journal of Nursing Research, 24,* 406–421.

Ellis, P. M. (2000). Attitudes toward and participation in randomized clinical trials in oncology: A review of the literature. *Annals of Oncology, 11,* 939–945.

Horne, R. (1999). Patients' beliefs about treatment: The hidden determinant of treatment outcome? *Journal of Psychosomatic Research, 47*(6), 491–495.

Jenkins, V., & Fallowfield, L. (2000). Reasons for accepting or declining to participate in randomized clinical trials for cancer therapy. *British Journal of Cancer, 82*(11), 1783–1788.

King, M. (2000). The effects of patients' and practitioners' preferences on randomized clinical trials. *Palliative Medicine, 14,* 539–542.

Klaber-Moffett, J., Torgerson, D., Bell-Syer, S., Jackson, D., Llewlyn-Phillips, H., Farrin, A., et al. (1999). Randomized controlled trial of exercise for low back pain: Clinical outcomes, costs, and preferences. *British Medical Journal, 319*(7205), 279–283.

Lambert, M. F., & Wood, J. (2000). Incorporating patient preferences into randomized trials. *Journal of Clinical Epidemiology, 53,* 163–166.

Larrabee, J. H., & Bolden, L. V. (2001). Defining patient-perceived quality of nursing care. *Journal of Nursing Care Quality, 16*(1), 34–60.

Lauver, D. R., Ward, S. E., Heidrich, S. M., Keller, M. L., Bowers, B. J., Brennan, P. F., et al. (2002). Patient-centered interventions. *Research in Nursing and Health, 25,* 246–255.

Lipsey, M. W. (1990). Design sensitivity: Statistical power for experimental research. Newbury Park, CA: Sage.

McCormack, B. (2003). Researching nursing practice: Does person-centredness matter? *Nursing Philosophy, 4,* 179–188.

McPherson, K., & Britton, A. (2001). Preferences and understanding their effects on health. *Quality in Health Care, 10*(Suppl 1), 161–166.

McPherson, K., & Chalmers, I. (1998). Incorporating patient preferences into clinical trials: Information about patients' preference must be obtained first. *British Medical Journal, 317*(7150), 78.

Morin, C. M., Gaulier, B., Barry, T., & Kowatch, R. A. (1992). Patients' acceptance of psychological and pharmacological therapies for insomnia. *Sleep, 15,* 302–305.

North-West Uro-Oncology Group. (2002). A preliminary report on a patient-preference study to compare treatment options in early prostate cancer. *BJU International, 90,* 253–256.

O'Connor, A. M., Mulley, A. G., & Wennberg, J. E. Jr. (2003). Standard consultations are not enough to ensure decision quality regarding preference-sensitive options. *Journal of the National Cancer Institute, 95*(8), 250.

O'Reilly, P. H., Martin, L., & Collins, G. (1999). Controversy in managing patients with prostate cancer. *British Medical Journal, 318,* 126.

Ruland, C. M. (1998). Improving patient outcomes by including patient preferences in nursing care. *Journal of the American Medical Informatics Association,* Suppl S, 448–454.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental design for generalized causal inference.* Boston: Houghton-Mifflin.

Sidani, S., & Braden, C. J. (1998). *Evaluating nursing interventions: A theory-driven approach*. Thousand Oaks, CA: Sage.

Streiner, D. L. (2002). The 2 "Es" of research: Efficacy and effectiveness trials. *Canadian Journal of Psychiatry, 47*(6), 552–556.

TenHave, T. R., Coyne, J., Salzer, M., & Katz, I. (2003). Research to improve the quality of care for depression: Alternatives to the simple randomized clinical trial. *General Hospital Psychiatry, 25,* 115–123.

Thomas, E., Craft, P. R., Paterson, S. M., Dziedzic, K., & Hay, E. M. (2004). What influences participants' treatment preference and can it influence outcome? Results from a primary care-based ransomised trial for shoulder pain. *British Journal of General Practice, 54,* 93–96.

Torgerson, D. J., Klaber-Moffett, J., & Russell, I. T. (1996). Patient preferences in randomized trials: Threat or opportunity? *Journal of Health Services Research Policy, 1*(4), 194–197.

Torgerson, D. J., & Sibbald, B. (1998). What is a patient preference trial? *British Medical Journal, 316*(7128), 360.

Tunis, S. R., Stryer, D. B., & Clancy, C. M. (2003). Practical clinical trials increasing the value of clinical research for decision making in clinical and health policy. *Journal of the American Medical Association, 290*(12), 1624–1632.

Vincent, N., & Lionberg, C. (2001). Treatment preference and patient satisfaction in chronic insomnia. *Sleep, 24*(4), 411–417.

Walter, J., Vincent, N., Furer, P., Cox, B., & Kjernisted, K. (1999). Treatment preference in hypochondriasis. *Journal of Behavioral Therapy and Experimental Psychiatry, 30,* 251–258.

Whittemore, R., & Grey, M. (2002). The systematic development of nursing interventions. *Journal of Nursing Scholarship, 34*(2), 115–120.

Author's Note

This study was supported by grant #NR05075 from the National Institutes of Health/National Institute of Nursing Research. The principal investigator was Dr. Souraya Sidani.

Comments or inquiries may be directed to Joyal Miranda, Faculty of Nursing, University of Toronto, 50 St. George Street, Toronto, Ontario M5S 3H4 Canada. Telephone: 416-946-8183. Fax: 416-978-8222. E-mail: joyal.miranda@ utoronto.ca

---

Joyal Miranda, RN, MN, is a PhD student in the Faculty of Nursing, University of Toronto, Ontario, Canada.

CJNR 2004, Vol. 36 N° 3, 116–129

Résumé

L'amélioration des soins axés sur les patients par le biais du développement des connaissances

Richard W. Redman et Mary R. Lynn

La demande pour des services de soins axés sur les patients met en évidence le besoin de développer les connaissances autant dans le domaine conceptuel qu'empirique. Les définitions et les éléments opérationnels des soins axés sur les patients comportent diverses questions conceptuelles. La satisfaction des désirs, préférences et attentes des patients constitue un élément commun à toutes les définitions. Dans le domaine de la recherche, les études portant sur les interventions comportent des défis sur les plans de la conception et du mesurage. Le développement d'interventions axées sur les patients ou conçues sur mesure selon les caractéristiques liées aux patients et l'environnement dans lequel ces actions seront menées figurent parmi ces défis. En se penchant sur ces questions vitales, la profession infirmière peut jouer un rôle clé dans l'évolution de la science de l'intervention et des connaissances dans le domaine des soins axés sur les patients.

Mots clés : soins axés sur les patients, interventions axées sur les patients

CJNR 2004, Vol. 36 N° 3, 116–129

Advancing Patient-Centred Care Through Knowledge Development

Richard W. Redman and Mary R. Lynn

The call for health-care services that are patient-centred raises the need for knowledge development in both the conceptual and empirical domains. The definitions and operational elements of patient-centred care present a variety of conceptual issues. A common element in all definitions is accommodation of patient wants, preferences, and expectations. In the research domain, intervention studies face both design and measurement challenges. These include the development of interventions that are patient-centred or tailored for both patient characteristics and the environment in which they will be delivered. By addressing these critical issues, nursing can play a key role in advancing intervention science and knowledge development in the domain of patient-centred care.

Keywords: patient-centred care, patient-centred interventions, organizational capacity

Introduction

A recent report by the Institute of Medicine (Committee on Quality Health Care in America, 2001) on changes needed in the delivery of health services delineates six features in need of reform: safety, efficiency, timeliness, effectiveness, equity, and patient-centredness. These improvements require the generation of specific knowledge about "best practices" or optimum means by which each can be achieved. While they are essential for the enhancement of health care, the notion of care being patient-centred presents two central challenges. The first relates to the discrepancy between the conceptualization and implementation of patient-centred care. Most providers would propose that their care is always patient-centred, yet patients might disagree. The second challenge relates to the methodology for developing and testing patient-centred interventions. Research designs and methods are generally standardized and participants are often unable to express their preferences or goals, as would be expected with a patient-centred approach to care.

Patient-centred care comes about as a result of the partnership that develops between the patient and the provider to ensure that care is based on joint decision-making, fostered by adequate education and support of the patient in the decision-making role. Such care is tailored to the patient's unique needs and preferences. Beyond the obvious oblig-

ation of nurses to know their patients' needs and preferences when planning care, the importance of patient-centred care is clearly demonstrated by the finding that patients who are centrally involved in the decision-making about their care have better outcomes than patients who are not (Holman & Lorig, 2000). However, despite the inherent appeal and obvious benefits of patient-centred care, little progress has been made in identifying ways of incorporating patient-centred concepts into health care in order to enhance the quality of care.

Several issues may be operational in the sluggishness of the movement towards patient-centred care. One issue is the lack of conceptual clarity and the lack of an operational definition of patient-centred care. Another emanates from the general lack of patient-centredness in research methods. In traditional intervention study designs, the intervention is not tailored to participant characteristics, needs, or preferences so as to be truly patient-centred. Additionally, intervention studies often do not take into account real-world contextual factors that affect the implementation of the intervention. The lack of patient-centredness in research results in a reluctance on the part of clinicians to integrate new interventions into their practice.

The development of knowledge that facilitates the delivery of patient-centred care requires clarification of the fundamental processes and the variables underlying such care as well as innovative research strategies that enable the investigation of patient-centred interventions. In this paper we will attempt to demonstrate that consideration of these factors will serve to promote patient-centred practice.

Conceptual Issues

Patient-centred care, commonly expressed as treatment of the patient as a unique person, actualizes a core value of nursing — individualization of care. This requires an understanding of specific patient needs and perceptions and, based on that understanding, the selection of optimal interventions to meet those needs. From this perspective, individualization of care is central in clinical decision-making (Radwin, 1996). Related to patient-centred care is the notion of patient participation in care. A concept analysis of patient participation has identified four essential attributes: a relationship between nurse and patient; a surrendering by the nurse of some degree of control or power; engagement, on the part of both nurse and patient, in intellectual and/or physical activities; and a positive outcome of participation (Cahill, 1996).

The benefits of patient-centred care include improvements in the patient's autonomy, functional status, quality of life, continuity of care, and health promotion behaviour (Committee on Quality Health Care in

America, 2001). Researchers, however, are not in agreement about how patient-centred care should be defined and measured. Furthermore, when researchers have attempted to measure patient-centred care, they seem to have defined it primarily from the perspective of the provider rather than that of the patient (Suhonen, Valimaki, & Leino-Kilpi, 2002).

Lutz and Bowers (2000) describe multiple perspectives on the interpretation and implementation of patient-centred care. Using concept analysis to examine these multiple perspectives, they found that the literature focuses on patient-centred care as designed in one of two ways in order to "meet the patient's needs": care organized around the patient's needs, or use of one's understanding of the patient's needs as a framework for care. Despite this focus on the patient's needs as central to the design of care, Lutz and Bowers found that patient-centred care was nominal only and was frequently defined and implemented from a traditional provider-centred approach, often resulting in delivery and outcomes not congruent with patient preferences. Varying definitions and conceptual views notwithstanding, an underlying theme has been identified: a fundamental concern with meeting patients' needs, wants, and/or expectations by respecting and integrating individual differences when delivering care (Lauver et al., 2002). Nursing, with its longstanding commitment to patient-centred care, is in the best position to lead conceptual and research efforts to develop interventions and models of care that incorporate patients' needs and preferences (Lutz & Bowers).

Issues of Research Design and Methods

One critical issue in the lack of patient-centredness in research is the prevailing view that the randomized clinical trial (RCT) design, with its standardization of treatment, application to all subjects, and random assignment to control and experimental groups, is the only acceptable way to test interventions. In the RCT model, the evaluation of interventions has two standard phases. In the efficacy phase the intervention is examined under tightly controlled conditions. Once efficacy has been established, in the effectiveness phase the intervention's performance is examined in the practice setting, under different conditions (Sidani, 1998). The RCT rigid design presents many challenges for intervention testing, particularly in the effectiveness phase.

The problems associated with the lack of utility of many intervention research findings and the fact that they are not always embraced in clinical practice have been well documented (Conn, Rantz, Wipke-Tevis, & Maas, 2001; Gross & Fogg, 2001; Sidani, Epstein, & Moritz, 2003). Research suggests that clinicians often do not change their clinical practice as new evidence becomes available because they do not agree with

that evidence or do not believe it will lead to improved outcomes in their patients (Cabana et al., 1999; Estabrooks, Floyd, Scott-Findlay, O'Leary, & Gushta, 2003). Clinicians' scepticism about and hesitancy to incorporate new evidence into their practice may stem from a belief that the evidence is not "real world" or relevant for their practice, where the focus is on tailoring care to patients' needs.

One approach to the promotion of patient-centred care is to test interventions that are patient-centred or tailored to clinically important patient characteristics. In the patient-centred intervention approach, methodological concerns focus on "realistic evaluation of the intervention consistent with nursing perspective underlying patient-centred care that is reflective of the natural conditions of everyday practice" (Sidani et al., 2003, p. 248). The intervention is tailored, and methodological modifications are made in the areas of selection of participants, treatment assignment, choice of measures, and extraneous factors — specifically, characteristics of the participant, the intervener, and the environment. The modifications are made with a view to enhancing the patient-centredness of research.

Tailored Interventions

In patient-centred care, interventions are tailored to the patient's specific capabilities, vulnerabilities, and, to the extent possible, preferences (Azar, 1999; Coward, 2002; Lauver et al., 2002; Sidani et al., 2003). Tailoring allows for efficient use of resources and for more specific treatment and more carefully controlled delivery of treatment. Tailoring also reduces the number of treatment dropouts that result when care providers and patients have different goals (Prochaska & DiClemente, 1983).

Historically, intervention studies are based on standardized interventions, assuming that "one size fits all." This approach underestimates individual variations in dose requirement and treatment acceptance, which may be one reason why some treatments are considered ineffective from the perspective of both the provider and the recipient. If a standardized intervention runs counter to patients' lifestyles, cultures, beliefs, or resources, it will be met by non-compliance or non-adherence, which in turn will influence outcomes. Therefore, the extent to which interventions can be tailored to the needs of participants, both the patient and intervention science will benefit.

Assignment to treatment. The design of an intervention study is generally based on random allocation of purportedly homogeneous participants to an experimental or control group. An alternative approach, one that is consistent with patient-centredness, incorporates input from participants before and during implementation of the intervention. According to the theory of psychological resistance (Brehm, 1966), when

people are placed in a position of perceived lack of control (eliminated freedom), as in the case of random assignment to a study group, the ensuing reactant behaviour can result in their quitting the study or investing minimally in their participation. Assignment to a specific treatment arm should be considered in the design of a patient-centred intervention evaluation study, as is done in partial RCTs. Participants often want to choose which arm of the study they are assigned to (Nield-Anderson, Dixon, & Lee, 1999). In a partial RCT (Bradley, 1993; Coward, 2002), participants with no treatment preference are randomized to an intervention arm and those with a preference are assigned to their preferred treatment. This approach reduces reactance behaviour and increases treatment compliance (Bradley). Additionally, it allows for examination of the effects of treatment preference on outcomes (Corrigan & Salzer, 2003; Gross & Fogg, 2001).

The effectiveness of tailored interventions depends in part on the degree to which the patients' situation and preferences can be considered in the selection of treatment. This matching of patients' situation and preferences to treatment arm requires a systematic allocation procedure based on a comprehensive evaluation of participants' characteristics and needs and explicit guidelines that link assessment results to specific treatment strategies. Ideally, patient-treatment matching guidelines are theoretically supported and empirically justified (Del Boca & Mattson, 1994).

Participant selection. In the most liberal patient-centred approach to research, participants are selected not on the basis of an exhaustive set of inclusion or exclusion criteria but primarily on the basis of whether they have the specific condition addressed by the intervention and whether they have characteristics that would make them resistant to the intervention. The self-selection bias is tracked, in order to provide information on how to appeal to those who did not enrol in the intervention (Sidani et al., 2003).

Choice of measures. In the patient-centred paradigm of research, instruments, or measures, are chosen not only for their traditional merits — reliability, validity, sensitivity to change — but also for their usability in routine clinical practice. Specifically, measures are selected if they are acceptable to the participants, easy to administer, and easy to score (Sidani et al., 2003). They have to be responsive to change generally and to clinically significant change(s) specifically (Deyo, Diehr, & Patrick, 1991; Guyatt, Deyo, Charlson, Levine, & Mitchell, 1989; Stewart & Archbold, 1993). These aspects are particularly important in the context of eventual translation and practical evaluation of the intervention.

Another issue related to instrumentation is the degree to which standardized instruments capture factors that reflect patient values or preferences regarding care and the outcomes of that care. This is particularly

important in intervention studies in which measures of health-related quality of life and functional status are included to supplement clinical or biological measures to assess the effectiveness of interventions. Generally, such measures are viewed as a way to include the patient's perspective on the condition or treatment and thus are described as patient-centred. However, many measures of health-related quality of life and functional status have been developed by providers, based on standard models, with little or no input from patients (Carr & Higginson, 2001). This raises the question of whether they are describing the patient's health from the perspective of the patient or from the perspective of the provider or society.

If a measure does not capture changes in health status or quality of life in terms that are meaningful to patients, then the responses of patients who are subsequently asked how they feel or to comment on their health status may not relate to changes in the measured health status, because the patient and the measure are not judging "change" on the same basis. If a measure is not patient-centred in terms of content or weight given to items or scales to reflect patients' values and preferences, and if it is used to assess the effectiveness of an intervention, the results are not likely to provide the kind of practical information that clinicians need.

Extraneous Influences

Extraneous factors such as the characteristics of the patient, intervener, and/or environment can directly affect or moderate outcomes (Lipsey, 1993). Patient characteristics that can influence outcomes include personal ones such as age, gender, education, and ethnicity, as well as health-related ones such as health status, comorbidities, severity of condition, and functional status. A social focus might also be included, in terms of resource availability, social support, or employment status. Characteristics of the intervener that must be considered include personal ones such as age, gender, ethnicity, and presentation and communication skills, and professional ones such as education, job satisfaction, and intervention skills (Epstein, 1995). Physical and psychosocial features of the environment can enhance or mitigate the effects of an intervention (Conrad & Conrad, 1994). Physical features of the environment include "comfort aspects" — that is, noise, light, temperature, familiarity, and overall appeal. Psychosocial features include the geographic, social, and cultural context of the study. The importance of each of these factors varies from study to study, but each can have direct and indirect effects on the outcomes of the intervention — although for the most part they may not be controllable.

In standardized approaches to evaluating interventions, the influence of extraneous factors is generally treated as "noise" and randomly distributed across treatment arms so that, while potentially adding to error vari-

ance, it is ignored. In the patient-centred approach to research, these factors are identified at the outset and considered in the analysis, to explore rival hypotheses as to the findings. Such considerations will add to the validity of the conclusions reached (Chen & Rossi, 1987; Cook & Campbell, 1979; Sidani & Braden, 1998).

These extraneous factors may be treated as moderators in the design and analysis of the intervention. Moderators influence the direction and/or magnitude of the effect of the independent variable on the dependent variable (Baron & Kenny, 1986; Kenny, 2003). Patient characteristics may moderate the intervention effect in addition to the moderation effect of most other extraneous factors noted earlier.

The day-to-day world of nursing is reflected in the design of patient-centred interventions, characterized by treatment tailored to patients' needs and preferences, and the design and conduct of intervention studies within the patient-centred approach to research. Nurses are trained to assess their patients before providing care. A natural extension of that process is assessment of the needs and wants of patients/participants in order to determine what will be done, when it will be done, where it will be done, and who will do it. We should not presume to "know" our patients well enough to design and implement an intervention that does not include their input and perspective; we should capitalize on the interactivity of nursing to advance our understanding of the responses of patients and participants to interventions they help design (Lynn, 1987). Clinicians frequently reject the findings from intervention studies as irrelevant to their clinical situation, often because they perceive the results as having limited applicability to the patients in their particular practice setting. The advent of the patient-centred intervention approach to knowledge generation provides a means by which real-world applicability of the findings can be built into the design and conduct of the study. While this requires some methodological trade-offs, these are outweighed by the potential gains to be made with regard to the clinical relevance of the findings.

Contextual Issues

Another challenge in developing and testing nursing interventions is assessing the capacity of the clinical setting and how this might influence the implementation of the intervention. Capacity issues include readiness of the organization to adopt change, skills of the practitioners, degree of flexibility in the systems of care, and availability of resources including the technological and clinical information systems needed to implement the intervention (Snyder-Halpern, 1999). Characteristics of the practice environment are not generally assessed when interventions are tested. Patient-centred care requires a provider who thinks critically and reflec-

tively. Factors that influence patient-centred practice include the nursing practice model, the skill mix, the adequacy of staffing on any given day, and the authority vested in the clinical decision-makers to implement patient-centred interventions (Brennan, 2002).

Naylor (2003) calls for increased attention to the context of care when interventions are being designed. When an intervention is being tested, it should be evaluated from the perspective of the patients receiving it as well as the clinical environment in which it will be implemented. Identification of effective strategies to assess the organizational environment, and the building of those strategies into intervention studies, might increase the probability of promising interventions being successfully integrated into clinicians' practice.

Recommendations

Given the increased dialogue and recognition of the importance of integrating patient-centred care into both practice and research, we offer several recommendations to guide knowledge development in this domain. Nursing is well positioned to address this challenge due to its philosophic commitment to patient-centred concepts and the nature of our research questions around the needs of individuals in health and illness.

Further Conceptual Development of Patient-Centred Care

The literature supports the need for refinement of the conceptual and operational aspects of patient-centred care (Lauver et al., 2002; Lutz & Bowers, 2000; Radwin, 1996). We lack clarification on the degree to which an individual's goals or priorities are, or should be, solicited by the provider so that they can be incorporated into the care plan. In addition, we lack guidelines for consistency in the solicitation of client input and in provider/client collaboration regarding patient-centred care. These components of patient-centredness will likely vary with the patient's health status and/or health-care choices. Further work is needed in this area so that patient-centred principles can be incorporated into both research and practice.

Integration of Patient-Centred Models in Intervention Research

Several models are available to guide the research community in the use of methods that incorporate components of patient-centred care into intervention research. In many ways, the arguments against the use of patient-centred techniques in experimental designs are generational: researchers trained in methods using the rigid RCT model as the gold standard view this model as the only acceptable one for testing and eval-

uating interventions, while those currently being trained are exposed to the notion that the lack of real-world application of RCT may make it a suboptimal choice when the goal is clinical application of the research findings. A patient-centred orientation in intervention research, particularly in effectiveness research, will facilitate the movement towards patient-centred care and thus lead to improved care. The research community should carefully consider the methodologic alternatives to prevailing models.

Sidani et al. (2003) describe a theory-driven approach to the evaluation of interventions. In this approach, the theory or conceptual model drives the selection of variables and the design of an intervention that incorporates the testing of the impact of selected patient-centred factors, such as patient characteristics, on the outcomes rather than controlling for their effects through design or randomization. In the theory-driven perspective, patient-centred principles can be incorporated into the design and implementation of the intervention, assignment of participants to treatment options, and selection of outcome measures.

In another approach to the integration of a patient-centred emphasis in intervention design, the research participants take part in evaluating the dosage of the intervention received so as to ensure accurate measurement. Sidani (1998) proposes a continuous scoring scheme for quantifying intervention dosage whereby patients are engaged in recording the treatment they receive in terms of time increments, frequency, intervals, or whatever calibration may be appropriate. This approach could be extended to include participants' involvement in the design and delivery of the intervention to incorporate aspects that are meaningful from their perspective.

Measuring treatment outcomes in patient terms, though not a part of most research studies, could provide insight into what indicators to use in evaluating the effectiveness of interventions. Many clinical scoring systems address general and condition-specific indicators from the clinician's perspective, and, when the results have been compared with patients' indicators of treatment effectiveness, discrepancies have been identified in the evaluations of clinicians and patients (Bayley, London, Grunkemeier, & Lansky, 1995). Incorporating the patient's views with regard to treatment effectiveness would enhance sensitivity when pre- to post-treatment changes are being evaluated.

Gross and Fogg (2001) discuss people's growing concerns about participating in RCTs and their unwillingness to comply with protocols that do not meet their needs. These authors stress the importance of identifying outcomes that are relevant for participants. Engaging participants as knowledgeable research partners would increase the patient-centredness of the study and the utility of its results. One way to increase

participant involvement in both protocol development and outcome measures is to have a participant advisory board assist with study and intervention design and selection of outcome measures. Such a board could provide insights that are otherwise unavailable to the investigators. Alternatives to randomization could also be considered, with a view to enhancing the patient-centredness of interventions. One technique would be to allow participants to choose their group assignment. While alternatives have disadvantages, they merit further examination as a means of building patient-centred aspects into research design and into the testing of interventions.

Evaluation of Available Measures and/or Development of New Measures

Carr and Higginson (2001) identify issues that are central to many standardized measures of health-related quality of life. Frequently these measures do not account for individual values and preferences, the cultural dimension, or the specific values, or weight, that individuals may assign to specific dimensions. Often they are designed from the perspective of researchers or health professionals, who assume they know what factors are relevant instead of asking patients or participants what aspects of their lives are important to them. The broader question is how well existing measures address issues of importance to patients in terms of determining the effects of treatment on quality of life or functional status. If a measure does not capture aspects of quality of life from the perspective of individual patients, it may not be sensitive to changes post-intervention or post-treatment because it is not measuring what is important to patients.

The development of individualized measures has so far been limited. Existing measures need to be refined and new approaches developed to capture dimensions that reflect the values of individual patients and research participants. In an individualized quality-of-life questionnaire developed by Bernheim (1999), patients record the specific areas in their lives that are most important to them and rate the current status of each using visual analog scales. This patient-centred approach is likely be sensitive to pre- and post-measurement of treatment effects in areas that are important to the patient.

Evaluation of existing instruments, particularly in the area of health-related quality of life, is essential. Revision of existing measures, or the development of new measures that are patient-centred, is an essential next step in adopting a patient-centred approach to the evaluation of treatments and interventions.

The challenges facing a patient-centred orientation in health services research are formidable. Nursing is well-positioned to lead this effort, which will require collaboration among researchers, clinicians, and

patients. A commitment to patient-centred intervention and research design holds promise for the advancement of knowledge development, nursing practice, and patient and organizational outcomes. Engaging participants as knowledgeable partners in research will increase the likelihood of our research endeavours having real meaning for both patients and clinicians.

References

Azar, B. (1999). New pieces filling in addiction puzzle. *APA Monitor, 30*(1), 12, 15.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Bayley, K. B., London, M. R., Grunkemeier, G. L., & Lansky, D. J. (1995). Measuring the success of treatment in patient terms. *Medical Care, 33*(4), AS226–AS235.

Bernheim, J. L. (1999). How to get serious answers to the serious question "How have you been?" Subjective quality of life (QOL) as an individual experiential emergent construct. *Bioethics, 13,* 272–287.

Bradley, C. (1993). Clinical trials — time for a paradigm shift? *Diabetic Medicine, 5,* 107–109.

Brehm, J. W. (1966). *A theory of psychological reactance.* New York: Academic Press.

Brennan, P. F. (2002). Patient-centered interventions in the context of the existing health care delivery system. Retrieved July 17, 2003, from http://stti.confex.com/stti/sos13/techprogram/paper_11652.htm

Cabana, M. D., Rand, C. S., Powe, N. R., Wu, A. W., Wilson, M. H., Abboud, P. A., & Rubin, H. R. (1999). Why don't physicians follow clinical practice guidelines? A framework for improvement. *Journal of the American Medical Association, 282*(15), 1458–1465.

Cahill, J. (1996). Patient participation: A concept analysis. *Journal of Advanced Nursing, 24*(3), 561–571.

Carr, A. J., & Higginson, I. J. (2001). Are quality of life measures patient centred? *British Medical Journal, 322*(June 2), 1357–1360.

Chen, H., & Rossi, P. H. (1987). The theory-driven approach to validity. *Evaluation and Program Planning, 10,* 95–103.

Committee on Quality Health Care in America. (2001). *Crossing the quality chasm: A new health system for the 21st century.* Institute of Medicine. Washington: National Academy Press.

Conn, V. S., Rantz, M. J., Wipke-Tevis, D. D., & Maas, M. L. (2001). Designing effective nursing interventions. *Research in Nursing and Health, 24,* 433–442.

Conrad, K. J., & Conrad, K. M. (1994). Reassessing validity threats in experiments: Focus on construct validity. *New Directions for Program Evaluation, 63,* 5–26.

Cook, T. J., & Campbell, D. T. (1979). *Quasi experimentation: Design and analysis issues for field settings.* Boston: Houghton-Mifflin.

Corrigan, P. W., & Salzer, M. S. (2003). The conflict between random assignment and treatment preference: Implications for internal validity. *Evaluation and Program Planning, 26,* 109–121.

Coward, D. D. (2002). Partial randomization in a support group intervention study. *Western Journal of Nursing Research, 24*(4), 406–421.

Del Boca, F. K., & Mattson, M. E. (1994). Developments in alcoholism treatment research: Patient-treatment matching. *Alcoholism, 11*(6), 471–475.

Deyo, R. A., Diehr, P., & Patrick, D. L. (1991). Reproducibility and responsiveness of health status measures: Statistics and strategies for evaluation. *Control Clinical Trials, 12*(4), 142S–158S.

Epstein, D. (1995). Characteristics of the intervener. *Proceedings of the Western Institute of Nursing Research Conference, 28,* 316.

Estabrooks, C. A., Floyd, J. A., Scott-Findlay, S., O'Leary, K. A., & Gushta, M. (2003). Individual determinants of research utilization: A systematic review. *Journal of Advanced Nursing, 43*(5), 506–520.

Gross, D., & Fogg, L. (2001). Clinical trials in the 21st century: The case for participant-centered research. *Research in Nursing and Health, 24,* 530–539.

Guyatt, G. H., Deyo, R. A., Charlson, M., Levine, M. N., & Mitchell, A. (1989). Responsiveness and validity in health status measurement: A clarification. *Journal of Clinical Epidemiology, 42*(5), 403–408.

Holman, H., & Lorig, K. (2000). Patients as partners in managing chronic disease. Partnership is a prerequisite for effective and efficient health care. *British Medical Journal, 320*(7234), 526–527.

Kenny, D. A. (2003). *Mediation.* Retrieved July 18, 2003, from http://users.rcn.com/dakeny/mediate.htm

Lauver, D. R., Ward, S. E., Heidrich, S. M., Keller M. L., Bowers, B. J., Brennan, P. F., et al. (2002). Patient-centered interventions. *Research in Nursing and Health, 25,* 246–255.

Lipsey, M. W. (1993). *Design sensitivity statistical power for experimental research.* Newbury Park, CA: Sage.

Lutz, B. J., & Bowers, B. J. (2000). Patient-centered care: Understanding its interpretation and implementation in health care. *Scholarly Inquiry for Nursing Practice: An International Journal, 14*(2), 165–187.

Lynn, M. R. (1987). Toward more credibility in nursing research: Researcher-subject collaboration. *Journal of Pediatric Nursing, 2,* 205–207.

Naylor, M. D. (2003). Nursing intervention research and quality of care. *Nursing Research, 52*(6), 380–385.

Nield-Anderson, L., Dixon, J. K., & Lee, K. (1999). Random assignment and patient choice in a study of alternative pain relief for sickle cell disease. *Western Journal of Nursing Research, 21*(2), 266–274.

Prochaska, J. O., & DiClemente, C. C. (1983). Stages and processes of self-change of smoking: Toward an integrative model of change. *Journal of Consulting Clinical Psychology, 51*(3), 390–395.

Radwin, L. E. (1996). "Knowing the patient": A review of research on an emerging concept. *Journal of Advanced Nursing, 23*(6), 1142–1146.

Sidani, S. (1998). Measuring the intervention in effectiveness research. *Western Journal of Nursing Research, 20*(5), 621–635.

Sidani, S., & Braden, C. J. (1998). *Evaluating nursing interventions.* Thousand Oaks, CA: Sage.

Sidani, S., Epstein, D. R., & Moritz, P. (2003). An alternative paradigm for clinical nursing research: An exemplar. *Research in Nursing and Health, 26,* 244–255.

Snyder-Halpern, R. (1999). Assessing health care setting readiness for point of care computerized clinical decision support system innovations. *Outcomes Management for Nursing Practice, 3*(3), 118–127.

Stewart, B. J., & Archbold, P. G. (1993). Nursing intervention studies require outcome measures that are sensitive to change: Part two. *Research in Nursing and Health, 16*(1), 77–81.

Suhonen, R., Valimaki, M., & Leino-Kilpi, H. (2002). "Individualised care" from patients', nurses' and relatives' perspective — a review of the literature. *International Journal of Nursing Studies, 39*(6), 645–654.

Authors' Note

Comments or inquiries may be directed to Richard W. Redman, CB# 7460, Carrington Hall, School of Nursing, University of North Carolina, Chapel Hill, North Carolina 27599-7460 USA. Telephone: 919-966-3734. Fax: 919-843-6212. E-mail: rredman@email.unc.edu

---

Richard W. Redman, PhD, is Associate Dean and Professor, and Mary R. Lynn, PhD, is Associate Professor, School of Nursing, University of North Carolina at Chapel Hill, USA.

CJNR 2004, Vol. 36 Nº 3, 131–145

Innovative Methodology

Personal Construct Theory: A Strategy for the Study of Multidimensional Phenomena in Nursing

Anne W. Snowdon

Nursing research is characterized by the study of complex phenomena relative to health behaviours, health-care services, illness, and hospitalization events. The challenge for researchers is to accurately capture and analyze multidimensional phenomena in the context of a dynamic interplay of events and interactions in clinical settings. Traditional methodologies measure a variety of concepts using strategies such as observation and questionnaires that rely on descriptive and inferential statistical analysis. However, the dynamic interplay of experiences, perceptions, and meanings in the social context of the clinical setting is more difficult to examine using these methodologies.

In a recent study, an innovative, multidimensional theory and accompanying methodology were employed to examine parents' experiences in the dynamic social context of the hospital setting. Personal Construct Theory, with its accompanying methodology, Repertory Grid Technique, is a new approach (based on an old theory) to nursing research that is especially well suited to the study of complex, multidimensional research questions. Implications for nursing research, theory development, and practice will be examined relative to the utility of Personal Construct Theory.

Underpinnings of Personal Construct Theory

Personal Construct Theory was first described by George Kelly in a two-volume work in 1955. Kelly's basic thesis was that people construct their own cognitive frameworks for anticipating and interpreting situations, persons, and events (Kelly, 1970). The formal structure of Personal Construct Theory comprises two fundamental tenets and 11 corollaries. It has been described as a metatheory, a theory about theories, that specifies the human process of living, describing the ways in which people anticipate events (Kelly, 1970).

The two tenets of the theory are constructive alternativism and "man-the-scientist" (Blowers & O'Connor, 1996). Constructive alternativism postulates that the universe is a domain open to continual revision, whereby people are constructivists who actively interpret and give meaning to the world around them. Thus, reality is assumed to be subjective, each of us interpreting the world according to our own meaning system with all of its possibilities. The act of construing the world never delivers "reality" to a person; rather, it delivers progressive approximations of reality based on the person's anticipations tested against the outcomes (Blowers & O'Connor). The metaphor of "theory" is used to describe networks of constructs through which people see and handle the universe of situations in which they participate (Fransella & Bannister, 1977).

Man-as-scientist delineates a systematic model of how individuals view, interpret, and know the world around them. It postulates that people act in the manner of scientists, formulating hypotheses, testing them against reality (i.e., against their previous attempts to know the world), and revising them if they turn out to be false or of limited use. Inherent in this tenet is the assumption that people have a need to anticipate and predict the future so they can learn and plan based on possible or expected outcomes (Kelly, 1955). Kelly believed that people strive to make sense of their universe, themselves, and the situations they encounter, and, in order to do so, develop and re-develop an implicit theoretical framework called a "personal construct system" (Fransella & Bannister, 1977). People are scientists who derive hypotheses (have expectations) from their theories (from personal constructs), subject these hypotheses to experiments (respond behaviourally and take active risks in terms of the experiments), observe the results of their experiments (live with the outcome of their behaviour), modify their theory (change their minds, change themselves, and grow), and the cycle continues (Fransella & Bannister). Man-as-scientist is the theoretical model of how people develop construct systems and how information is processed through personal construct systems in a dynamic and cyclical manner.

The corollaries of Personal Construct Theory, which illustrate its principles, are as follows: a person anticipates events by construing their replications (construction corollary); each person characteristically evolves, for the purpose of anticipating events, a construct system embracing ordinal relationships between constructs (organization corollary); a person's construct system consists of a finite number of dichotomous constructs (dichotomy corollary); each person chooses the alternative in a dichotomized construct through which he or she anticipates the greater possibility for the elaboration of his or her system (choice corol-

lary); a construct is convenient for a finite range of events (range corollary); each person's construct system varies as he or she successively construes the replications of events (experience corollary); the variation in a person's construct system is limited by the permeability of the constructs within whose range of convenience the variants lie (modulation corollary); a person may successively employ a variety of construct subsystems that are inferentially incompatible with each other (fragmentation corollary); to the extent that the construction experience of one person is similar to that of another, the two processes are psychologically similar (commonality corollary); to the extent that one person construes the construction processes of another, he or she may play a role in a social process involving the other person (sociality corollary). The corollaries of Personal Construct Theory describe the development and evolution of construct systems and their application to or imposition on the events and interactions experienced by individuals.

Constructs

According to Kelly (1955), individuals retain dimensions that are relevant for them in order to form impressions of people, objects, and events. This is the heart of the interpretive process, or the act of construing, which leads to the formation of constructs that are "transparent patterns or templates" of the realities that make up the world (Blowers & O'Connor, 1996). The construct is a consistent way for each individual to make sense of some aspect of reality in terms of similarities and differences among objects and events. It is a process of determining, using polarized dimensions called constructs, whether people, places, or things are similar or different (Blowers & O'Connor). Constructs locate an event, help the individual to understand it, and then anticipate it. They are imposed upon events, not abstracted from them (Kelly, 1955). When imposed, a construct is used to distinguish events and group them. Thus, a construct is the distinction that individuals make between the events they experience (Kelly, 1955). Each construct represents a pair of rival hypotheses, either one of which may be applied to a new situation that people seek to interpret according to their attitudes, beliefs, and experiences. By testing hypotheses to see which best fits his or her expectations, the person can retain them temporarily, revise them, or replace them (Blowers & O'Connor). Constructs are points of reference that a person projects upon an event in an effort to more fully understand it.

Construct *Defined in Relation to* Concept

The term *concept* is heavily used in both the social sciences and the health sciences, and it has numerous meanings. Kelly (1955) defines *construct* in

terms of what it is and how it is different from *concept*, in order to clearly illustrate a fundamental aspect of his theory. Specifically, a concept is a static property or characteristic of two or more objects that otherwise are distinct (Kelly, 1955). Kelly would argue that a concept is simply a characteristic property used to classify and define objects in terms of their similarities and differences and that a concept is a finite characterization of an object or event.

A construct can be defined as follows: the interpretation of the use of concepts, or the process of knowing or, more importantly, coming to know (Blowers & O'Connor, 1996); the process by which individuals ascribe meaning to and interpret people, places, and situations or events in the world around them; a basis for discrimination and association within a system that is fluid and evolves and changes as new events are experienced (Kelly, 1970); the contrast that a person perceives — it does not represent an object and is not categorical or concrete (Kelly, 1970).

Personal Construct Theory might at first appear to be similar to symbolic interactionism, since both theories focus on how people construct meaning as they come to know or make sense of the world around them. Both theories generally describe how individuals ascribe meaning and how their behaviour is linked to the meaning they attach to life experiences. Interestingly, symbolic interactionism has been used widely as a theoretical basis in nursing research, yet Personal Construct Theory has remained unnoticed.

There are significant differences in the two theories. In symbolic interactionism, two people agree on the meaning ascribed to things in their environment (Benzies & Allen, 2001); behaviour is an outcome of the meanings derived from interaction (Benzies & Allen); meanings emerge from the individual's interactions and from societal expectations through reciprocal interaction (Benzies & Allen). In Personal Construct Theory, the individual imposes meaning on the event/experience based on the construal process used to understand and distinguish it from previous experiences (Kelly, 1970). According to Kelly (1970), individuals may or may not have agreement between their personal construct systems, but they interact in order to hypothesize and test their construal of a situation or a role they assume.

The main advantage of Personal Construct Theory for nursing research is the well-defined methodology developed to examine individuals' personal construct systems and ways of knowing. Nurse researchers can potentially use Repertory Grid Technique, a methodology developed specifically from Personal Construct Theory, to examine the complexities of meaning systems and the multidimensional nature of health events and health experiences.

Methodology: Repertory Grid Technique

The repertory grid is not only a fertile instrument but a very flexible one that has been widely used in clinical (therapeutic) applications as well as non-clinical applications such as industrial training, quality control, and management development (Shaw, 1981). Behind each act of interpretation or judgement that people make (consciously or unconsciously) lies their personal belief system with regard to the meaning and significance of the event they are experiencing (Fransella & Bannister, 1977). The repertory grid is a way of exploring the structure and content of such implicit belief systems or theories (Fransella & Bannister).

A full repertory grid has three components: elements, which define the people and situations upon which the grid will be based; constructs, or the ways in which the subject groups and differentiates between the elements; and a linking mechanism (i.e., a ranking or rating scale), which indicates how each element is being construed or assessed using each construct (Adams-Webber, 1981). The elements determine the focus of the grid. They must be elicited and defined as specifically as possible (Adams-Webber), and they must be homogeneous, all drawn from the same category of interest to the researcher.

Repertory Grid Technique was used in a study examining mothers' experiences during their child's unexpected acute hospitalization. The hospitalization of a child has been well documented as a highly stressful experience for parents. However, the meaning(s) and multidimensional nature of this experience has not been examined. The following detailed account of how the repertory grid was used in this study will highlight its flexibility and the potential it offers to nurse researchers.

Development of the Repertory Grid

The focus of the study were the elements (people and situations) associated with the pediatric unit of a community general hospital. Such elements must be representative and cover the full range of elements associated with the area to be investigated (Adams-Weber, 1981). All mothers in the study had to be able to relate directly to the elements (i.e., likely to have experienced all the elements) in the grid (Adams-Weber). There can be as few as six elements in a grid typically used for non-clinical applications and as many as 25 elements in a grid used for research and clinical therapies (Adams-Webber).

The elements were elicited during early pilot testing using open-ended interviews with a group of mothers who were representative of the sample population (i.e., had experienced the hospitalization of a child). During the interview, the mothers described the people with

whom they interacted and the situations or events they experienced. The people they most commonly described were nurses, physicians, and other parents who were present during the child's hospitalization. After eight interviews, elements emerged that were common to the experiences of all the mothers. These were the elements placed in the grid.

A grid has been broadly described as a sorting task that produces primary data in a matrix form to allow for the assessment of relationships (Fransella & Bannister, 1977). Thus, elements are the objects of people's thoughts and constructs are the qualities that people attribute to these objects (Adams-Webber, 1981). Constructs for a research-focused grid are generated using a similar strategy of open-ended interviews. Although there are a variety of strategies for eliciting constructs, the classical approach is to use dyads: two elements are grouped together and the participant is asked to describe their similarities or differences (Adams-Webber). As the participants describe the differences between the two elements, they also describe the meaning they ascribe to each. This approach is used when the elements are complex, such as the relationship between a nurse and a parent in a hospital setting. The goal is to elicit the differentiation in meaning between the two elements. One should avoid focusing on the logical opposites but, rather, focus on the opposite meanings (Adams-Webber). In this study, the researcher selected two elements that were very different for the mothers, such as a nurse who was perceived as helpful and effective and a nurse who was perceived as unsupportive. The mothers were asked to "think of the nurse who you really enjoyed working with and compare him or her to the nurse you found difficult to work with." The dyad establishes a context of discrimination between two elements, so that differentiation between the two elements can be easily described and elicited. Constructs associated with the ineffective nurse included: "[doesn't] spend time with me," "controls my child's care," "not confident," "doesn't communicate with me." These constructs were opposite in meaning (not logically opposite) to those associated with the highly effective nurse: "meets my child's emotional needs," "flexible with rules and policies," "relinquishes control over care," "not intimidating." Verbal labels or adjectives that are logical opposites are more typical of semantic differential scaling. Grids can be used for semantic differential scaling, but in this case the constructs and bipolar adjectives would be developed by the researcher and placed in the grid using standardized verbal labels. In the present study, the verbal labels used by the mothers during the interviews were those used to identify the constructs in the grid. Thus, the uniqueness of the mothers' constructs was captured in their own words. Figure 1 illustrates the structure of the grid used for the mothers in the study.

Figure 1 ***Sample Page of Repertory Grid for Mothers***

Informed 1 Always 2 Moderately 3 A little				**Uninformed** 5 A little 6 Most of the time 7 Always		
1	2	3	Effective nurse	5	6	7
1	2	3	Ineffective nurse	5	6	7
1	2	3	Recovery room nurse	5	6	7
1	2	3	Doctor (Adm.)	5	6	7
1	2	3	Other parents	5	6	7
1	2	3	ER admission	5	6	7
1	2	3	Unit admission	5	6	7
1	2	3	Going to OR	5	6	7
1	2	3	In Recovery room	5	6	7
1	2	3	When child in pain	5	6	7
1	2	3	When mom leaves child	5	6	7
1	2	3	Mom stays at bedside	5	6	7
1	2	3	Decision re child's care	5	6	7
1	2	3	At discharge	5	6	7
1	2	3	IV procedure	5	6	7

For the mothers, the elements of primary importance were elements of the hospital system. These included the people with whom the mothers interacted (i.e., nurses and doctors) and the situations they experienced during the child's hospitalization (e.g., admission to hospital, making decisions about the child's care, procedures, admission to emergency room). An identical process of eliciting constructs and elements was used with nurses in the hospital setting in which the study was to take place. Constructs and elements that were relevant for nurses were elicited through open-ended preliminary interviews designed for repertory grid development.

The repertory grid matrix was administered using a structured interview when the mothers were asked to interpret each element according to each of the constructs using a rating scale. For example, the mothers were asked the following question: "Think back to when your child was

admitted to the hospital [element]. Did you feel 'informed' or 'not informed' [construct]?" They were then asked to rate the degree to which they felt informed (1 = very well informed, 2 = somewhat informed, 3 = a little informed or not informed, 5 = a little not informed, 6 = somewhat not informed, 7 = not at all informed). A score of 4 was a neutral score entered into the grid data by the researcher whenever a construct was not applicable (not experienced) for a particular mother. In order to ensure accuracy, the researcher administered the grid to each participant and circled the rating score for each element according to each construct. Any element that was not rated was left blank and received a score of 4 when data were entered into the Repgrid software for analysis (Shaw, 1996).

Theoretical Assumptions of Grids

All grids must consider the issue of range of convenience, which states: "A construct operates within a context and there are a finite number of elements to which it can be applied by a given person, at a given time" (Fransella & Bannister, 1977, p. 6). For a given act of construing at a given time, the range of convenience of a person's constructs is always limited (Fransella & Bannister). For example, the construct of "happy" can be applied to a number of people or situations but not to an inanimate object. Thus, the elements to which a construct would consistently be applied is limited. The principle of range of convenience must be carefully considered during both grid development and data analysis. For grid construction, Kelly (1955) derived a primary rule: "For given persons completing a grid, all elements must be within his/her range of convenience." Thus, the constructs in the grid for this study applied only to the elements (people and situations) in the hospital where the study was conducted and only to the mothers or nurses who were functioning there at the time of the study. For the purpose of ensuring that the elements were within the range of convenience, the mothers and nurses were asked to indicate when a construct was not applicable to an element. In addition, pilot testing of the grid was conducted with a small group of mothers and nurses to ensure that all the elements were familiar to the mothers and that the participants readily applied the constructs to the elements in the grid. Further, the elements had to be representative of the pool from which they were drawn (Fransella & Bannister). According to this principle, if the test is to indicate how the participant construes or understands other people or situations, then the other people or situations appearing as elements in the test must be sufficiently representative of all the people and situations with whom the participant interacts in the identified setting (Fransella & Bannister). The elements in the grid for this study included all the people and situations that the mothers most

often interacted with during their child's hospitalization and those that the nurses were most likely to interact with in their clinical practice with families in the hospital setting.

The second theoretical principle is the organization corollary, which postulates that each person characteristically evolves, for his convenience in anticipating events, a construction system embracing ordinal relationships between constructs (Fransella & Bannister, 1977). Specifically, when constructs and elements are elicited during open-ended interviews, the verbal labels the participants use in the grid-development phase are used to identify the constructs. Kelly (1955) asserts that the words that the client uses to describe meaning are the verbal labels employed for the constructs, rather than the dictionary or professional meanings that are typical of other methodologies. According to Personal Construct Theory, constructs are interrelated and, when further analyzed, can be grouped to reflect underlying dimensions or core constructs that are the central framework of an individual's personal construct system (Kelly, 1955). Each person has a small number of core constructs or dimensions in his or her personal construct system. The purpose of the grid analysis in the present study was to examine the hierarchical relationships between the elicited constructs and the core elements, in order to describe the underlying core dimensions of the mothers' and nurses' construct systems.

Analysis of Grid Data

According to Kelly, the conceptual grid is a premathematical representation of an individual's psychological space and is designed to set the stage for mathematical analysis of that space (Fransella & Bannister, 1977). Analysis is based on the assumption that statistical relationships within the grid represent relative stability and permanence in a person's construct system (Fransella & Bannister). Principal components analysis (PCA) was used to examine the numerical ratings that reflected patterns and relationships between the elements and constructs for both the mothers and the nurses. In PCA, the unit of analysis is the correlation coefficient. PCA transforms the raw data into a standardized set of correlations with a mean of zero and a standard deviation of one. A mean of zero means the data set has a common origin, which ensures that the differences between each rating are all standardized from a common mean of zero. The correlations are actually the correlations between entire *columns* of numbers (i.e., the rating of each element according to each construct) and entire *rows* of numbers (i.e., the rating of an element across every construct). The data output is captured in a conceptual map whereby the correlation coefficients are really the angle between two vectors mathematically. High correlations produce very small angles between two vectors, meaning that the subject views these two elements in a very

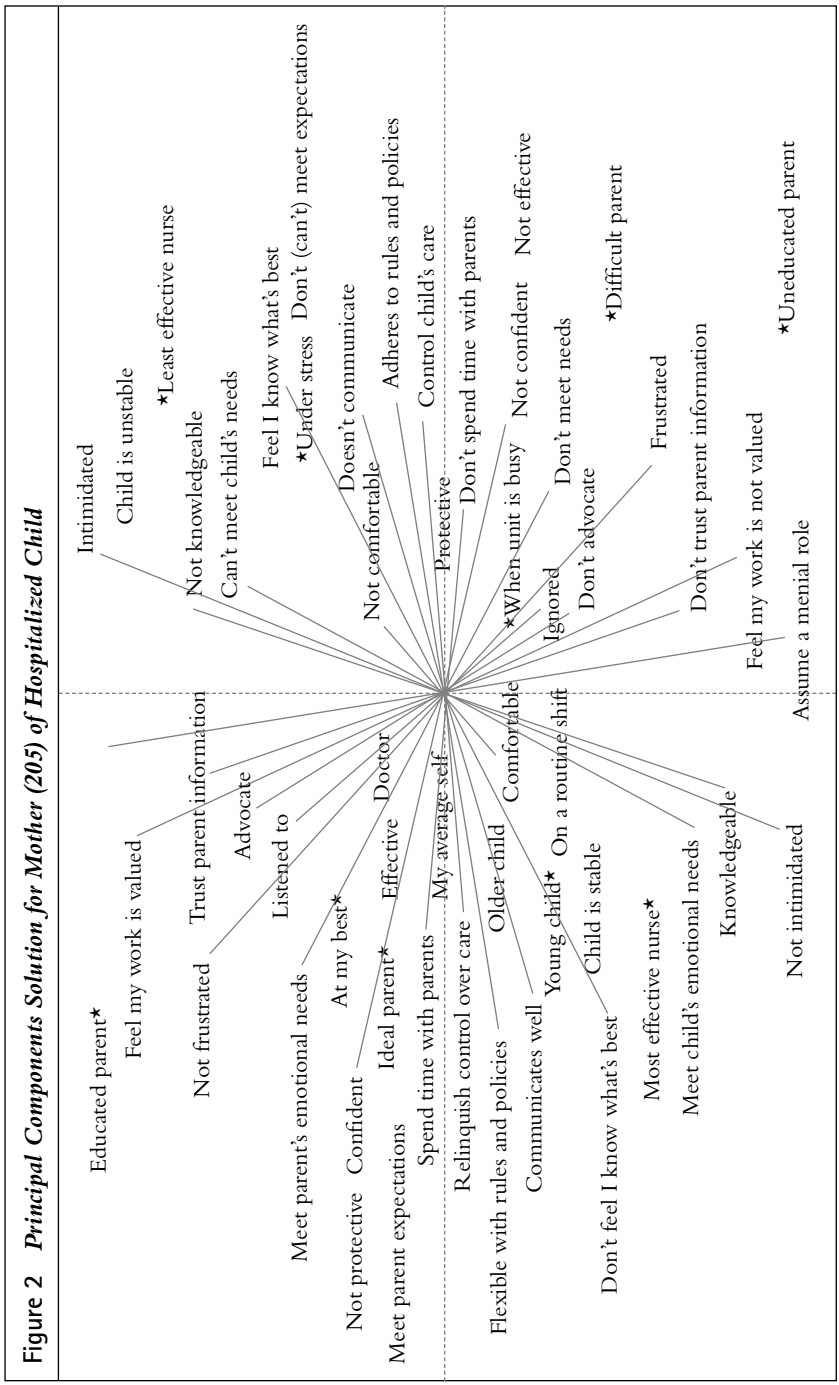


Figure 2 ***Principal Components Solution for Mother (205) of Hospitalized Child***

similar way, as they are highly correlated. When the correlations are low the participant views these two elements very differently, as the angle between the vector approaches 90 degrees. Figure 2 illustrates the vectors of the principal components solution for one of the mothers. Thus, principal components maps are simply vectors representing the correlations between the columns or rows of ratings on the repertory grid, and the angle between two elements represents the size of the correlation.

PCA is a data-reduction technique that draws a manufactured vector through as many vectors as possible to illustrate the underlying components in the data. The first line drawn through the vectors accounts for the maximum amount of variance in the data. The second line drawn through the vectors is as different as possible from the first and accounts for as much of the variance as possible once the first component is removed. A third component that is as different as possible from the first two is drawn through the vectors to account for as much of the remaining variance as possible when the first two components are removed. The first principal component is always 90 degrees to component two, which is 90 degrees to either of the first two components, implying that the components are not correlated.

PCA vector maps are mathematically separate for constructs and elements. However, because the data are standardized (have a common zero), a transformation matrix is used to multiply the element matrix by the transformed matrix, which has the effect of moving the elements matrix into the dimensional space of the construct principal components solution. What makes the innovation useful for nurse researchers is its ability to move the elements into the three-dimensional space of the construct matrices. The final solution is rotated using the Varimax rotation to further clarify the underlying construct dimensions in the data (Blowers & O'Connor, 1996). The data output includes a "raw grid" data display (Figure 3), a cluster analysis, and the principal components map (Princom map), which is generated for each participant (Figure 3). Cluster analysis was not used in this study since it does not offer the detailed mapping of the construct and element relationships in the data. The Princom map is a conceptual map illustrating the meaning that a mother or nurse attaches to the people or situations they encounter in the hospital setting. With each Princom map, the loadings for each element and construct are generated. Each loading represents the proximity of the construct or element to the principal component that accounts for the most variance possible. Higher loadings indicate a stronger relationship between the construct or element and the principal component. Analysis of the meaning of each of the underlying principal components in a solution is defined by the constructs that are most strongly loaded onto the principal component.

Figure 3 ***Data Output for Repertory Grid***

Display: Mother 203
Elements: 11; Constucts: 22; Range: 1 to 7; Context: hospital ward

		1	2	3	4	5	6	7	8	9	10	11		
Listened to	1	1	4	1	4	4	1	4	1	1	4	2	1	Not listened to
Adknowledged	2	1	4	1	4	4	1	4	1	1	1	6	2	Ignored
Respected	3	1	4	1	4	4	1	4	1	1	1	5	3	Not respected
Valued	4	1	4	1	4	4	1	4	1	1	1	6	4	Not valued
Caring	5	1	4	1	4	4	1	4	1	1	1	6	5	Not caring
Helpful	6	1	4	1	4	4	1	4	1	1	4	3	6	Not helpful
In control	7	1	4	1	4	4	1	4	1	1	1	7	7	Not in control
Included	8	1	4	1	4	4	1	4	1	1	1	7	8	Excluded
Involved	9	1	4	1	4	4	1	4	1	1	1	6	9	Not involved
Participates	10	1	4	1	4	4	1	4	4	1	1	3	10	No participation
Assume I don't know anything	11	3	4	4	4	4	4	4	4	4	4	4	11	Assume I know everything
No regret	12	1	4	1	4	4	1	4	1	1	1	7	12	Regretful
Not doubtful	13	1	4	1	4	4	1	4	1	1	1	5	13	Doubtful
Not fearful	14	1	4	1	4	4	1	4	1	1	1	6	14	Fearful
No guilt	15	1	4	1	4	4	1	4	5	1	1	5	15	Guilty
Manageable	16	2	4	1	4	4	1	4	1	1	1	5	16	Not manageable
Certainty	17	1	4	1	4	4	2	4	5	1	1	5	17	Uncertainty
Pleased	18	1	4	1	4	4	1	4	1	1	1	5	18	Disappointed
Satisfied	19	1	4	1	4	4	1	4	1	1	1	7	19	Dissatisfied
Reassured	20	1	4	1	4	4	1	4	1	1	1	5	20	Intimidated
Informed	21	1	4	1	4	4	1	4	1	1	1	7	21	Not informed
Understanding	22	1	4	1	4	4	1	4	1	1	1	5	22	Don't understand parents

1	2	3	4	5	6	7	8	9	10	11	
											11 During procedures
											10 Making decisions
											9 Mom stays at bedside
											8 When mom leaves child
											7 Child is in pain
											6 Unit admission
											5 Emergency admission
											4 Other parents
											3 Doctor
											2 Ineffective nurse
											1 Effective nurse

Overview of Findings

The purpose of this study was to conduct a multidimensional evaluation of how mothers perceive and ascribe meaning to their interactions in the hospital setting and how nurses perceive and ascribe meaning to their interactions with mothers. Mothers were approached 24 hours following the admission of their child to hospital and completed the repertory grid instrument during a structured interview. Nurses were interviewed using the repertory grid according to the convenience and availability of the nurse on the unit. Twenty mothers and 20 nurses participated. The hospitalization was the first for each family. The participating nurses were primarily Canadian citizens (65%) and had an average of 14 years of nursing experience.

The findings demonstrate that the mothers ascribed three underlying dimensions of meaning to their experience. In the first dimension they evaluated their interactions with nurses in the hospital setting, which were pivotal to their overall hospital experience, whether positive or negative. The second dimension was the mothers' ability to play their parental role in a meaningful way (i.e., to be involved in their child's care, informed, valued). The third dimension was the mothers' active involvement in their child's care, which was based not only on their sense of efficacy or perceived ability to become actively involved, but also on their decision to do so, which was influenced by their interactions with nurses. This was a powerful, emotional dimension: the decision to become involved in care resulted in significant emotional experiences. The mothers indicated that their most critical transactions in the hospital setting were mediated by frontline nursing staff, who played a central role in the meaning of the mothers' experiences.

In contrast, the nurses' meaning systems were very diverse and complex. There were more components to the nurses' PCA, with a more even distribution of variance accounted for by each of the dimensions. The nurses' first dimension was self-evaluation of their effectiveness as a professional, which amounted to a self-portrait of their performance in different circumstances and situations. This dimension concerned how the nurses felt (i.e., valued by parents, frustrated, comfortable, intimidated), what they were able to do for families (i.e., meet expectations and needs), and the strategies they used when interacting with mothers (i.e., communication, relinquishment or exertion of control). Control was a dominant theme in this component across all the nurses in the study. The second dimension was the nurses' ability to work with the mother or the situation, which was largely based on their assessment of the mother as "ideal" or "difficult," "educated" or "uneducated." When a mother was viewed as ideal, the nurse perceived an ability to work with her in an

effective manner. The third dimension was the nurses' ability to form a relationship with the mother or, alternatively, to limit their interactions with and control the behaviour of the mother. This dimension was, again, based on the nurses' evaluation of the mother; nurses formed relationships with mothers they perceived as ideal but limited their contact with those they perceived as difficult or "uneducated."

The findings indicate that there are clear differences in what is important to mothers and what is important to nurses in the hospital setting. The mothers valued their relationship with the nurse first and foremost, whereas nurses most valued their knowledge and competency in their professional role, using control as a strategy whenever they perceived a mother as less than ideal.

Conclusion

Personal Construct Theory and the accompanying Repertory Grid Technique offer a unique approach to the examination of personal construct systems. The outcome of this study was multidimensional mapping of the meanings of the experiences of both mothers and nurses in the hospital setting. Ultimately, Repertory Grid Technique was able to delineate the complex dynamics of mothers' interactions with nurses, something that empirical work has so far been unable to do. The flexibility of this technique allows nurse researchers to focus on any type of clinical phenomenon or health issue. Because of its multidimensionality, the technique presents a range of opportunities for descriptive research as well as intervention research using multidimensional mapping to inform nurse researchers about complex clinical phenomena and to show nurses the extraordinary impact they can have on the meanings of health events and health experiences in their clinical practice.

References

Adams-Webber, J. (1981). George A. Kelly as scientist-professional: An appreciation. In M. G. Shaw, *Recent advances in Personal Construct Technology* (pp. 1–7). New York: Academic.

Benzies, K. M., & Allen, M. N. (2001). Symbolic Interactionism as a theoretical perspective for multiple method research. *Journal of Advanced Nursing, 33*(4), 541–547.

Blowers, G. H., & O'Connor, K. P. (1996). *Personal Construct Psychology in the clinical context*. Ottawa: University of Ottawa Press.

Fransella, F., & Bannister, D. (1977). *A manual for Repertory Grid Technique*. London: Academic.

Kelly, G. A. (1955). *The psychology of personal constructs* (Vol. 1). New York: Norton.

Kelly, G. A. (1970). A brief introduction to Personal Construct Theory. In D. Bannister (Ed.), *Perspectives in Personal Construct Theory* (pp. 1– 30). London: Academic.

Shaw, M. G. (1981). *Recent advances in Personal Construct Technology*. New York: Academic.

Shaw, M. G. (1996). *RepGrid* (Version 2.1b). Calgary: Centre for Personal Computer Studies.

Author's Note

This study is based on the author's PhD dissertation, which was supported by Doctoral Fellowship #752-98-0115, Social Sciences and Humanities Research Council of Canada.

Anne W. Snowdon, RN, PhD, is Associate Professor, Faculty of Nursing, University of Windsor, Ontario, Canada.

CJNR 2004, Vol. 36 Nº 3, 147–151

Book Reviews

Action Research in Health

Ernie Stringer and William J. Genat
Upper Saddle River, NJ: Pearson Education, 2004. 204 pp.
ISBN 0-13-098578-3

Reviewed by Nazilla Khanlou

Action Research in Health is a new work by Stringer and Genat that will be of great value for students, educators, and practitioners of action research (AR). The book's intended audience is health professionals, with nurses being a central focus of discussion. In a reader-friendly format, the book addresses the why and how of AR in the health field. Recognizing the transformational value of research, the authors pay heed to both technical and human dimensions of AR throughout the volume.

The first two of nine chapters provide background information on AR, its purpose as a systematic and cyclical process of inquiry, and issues of underlying research paradigm (positivistic, naturalistic) and method (quantitative, qualitative). Stringer and Genat articulate their approach to AR as one that comprises participatory and interpretive features. In Chapter 3 they discuss preliminary steps in conducting AR. This chapter has several creative and unique aspects. For example, the human dimensions of AR are discussed under a section titled "With Head, Heart and Hand." In addition to standard criteria for assessing the trustworthiness of data, participatory and pragmatic validity are considered. An index of engagement considers issues of resistance, apathy, interest, and excitement by participants in AR.

The subsequent chapters are organized around the "Look-Think-Act" cycle of AR that is introduced in Chapter 3. The "Look" (information) feature concerns gathering information from stakeholders. To this end, Chapter 4 discusses in detail qualitative gathering procedures, emphasizing the interview as the starting point for inquiry. The "Think" (reflection) aspect, addressed in Chapter 5, relates to data analysis. Particular attention is given to analyzing epiphanies. The "Act" (action) feature distinguishes this book in terms of the breadth of information it provides on communication and planning activities as part of the research process. For example, Chapter 6 addresses the importance of communication in AR through such means as reports, presentations, and performances. Stringer and Genat argue that in AR reports one attempts to

"produce evocative accounts, conveying accurate insights into the impact of events on people's lives" (p. 119). Chapter 7 examines the relevance of AR for various health frameworks, including the nursing process, in improving professional practice. Because of the number of frameworks considered, the authors present parts of this chapter in a list format instead of as an in-depth discussion of the implications for practice. However, the particular relevance of AR to primary health care is insightful. The last two chapters, which could have been labelled appendices, provide a wealth of AR resources: Chapter 8 presents case studies in which AR has been used in various health contexts; Chapter 9 concludes the book with information on online resources on AR as well as related topics such as women's health, community/government/international organizations, and university programs.

Because Stringer and Genat consider AR an interpretive approach to qualitative inquiry, *Action Research in Health* provides a more in-depth discussion of naturalistic inquiry than of positivist research. However, this perspective is not shared by everyone. It can be argued that AR may entail qualitative, quantitative, or integrated approaches, depending on the focus of inquiry. What brings AR closer to the naturalist paradigm is (1) theoretical congruency between AR and critical inquiry, and (2) methodological congruency between AR's cyclical nature and emergent designs arising from qualitative research. In addition, missing from the discussions in the early chapters of the book is situating AR in relation to participatory research and participatory action research (see Brown & Tandon, 1983; Wallerstein & Duran, 2003).

Among the many strengths of this book is the humanist and passionate voice that the authors use in discussing the practice of AR. Through the use of box inserts, Stringer and Genat draw from their own AR experiences in various health contexts. The information is clearly presented throughout. Where relevant, figures and diagrams are used to augment the text. Methodological chapters begin with an overview of the concepts to be addressed and conclude with a summary of the content discussed. Other pedagogical strengths include the juxtaposition of the Look-Think-Act cycle and an AR Sequence conceptual framework and the substance of the various chapters.

Overall, *Action Research in Health* is a valuable book for diverse audiences. In education, it can be used by educators to teach a new generation of students about AR. In practice, the book provides detailed information on AR and interpretive qualitative methods for nurses to consider applying AR in various health-care settings.

References

Brown, L. D., & Tandon, R. (1983). Ideology and political economy in inquiry: Action research and participatory research. *Journal of Applied Behavioral Science, 19*(3), 277–294.

Wallerstein, N., & Duran, B. (2003). The conceptual, historical, and practice roots of community based participatory research and related participatory traditions. In M. Minkler & N. Wallerstein (Eds.), *Community-based participatory research for health* (pp. 27–52). San Francisco: Jossey-Bass.

Nazilla Khanlou, RN, PhD, is Assistant Professor, Faculty of Nursing and Department of Psychiatry, University of Toronto, Ontario, Canada. She is Course Coordinator for the undergraduate course Nursing Research and Scholarship and teaches its Participatory Action Research Stream.

Reading, Understanding, and Applying Nursing Research: A Text and Workbook, 2nd ed.

James A. Fain
Philadelphia: F. A. Davis, 2004. 439 pp.
ISBN 0-8036-1112-9

Reviewed by Joan Rosen Bloch

In *Reading, Understanding, and Applying Nursing Research: A Text and Workbook,* James A. Fain provides an interesting textbook for nursing research students at both the undergraduate and graduate levels.

Although easy to read, this work is comprehensive rather than a superficial survey of key research topics. Fain presents a creative collection of materials from a variety of sources that support the theme of his book: evaluating and integrating research findings into practice. The figures, excerpts from other documents, and summary tables are interesting, substantive, clinically relevant, and practical; therefore, it is applicable to nursing practice. For example, Table 3.2 (p. 51) succinctly lists the purposes of a literature review. Yet Table 3.3 (p. 52) gives lengthy, thorough instructions in conducting a search of the literature. Then, Fain discusses various sources of information (p. 53) and includes a small table of nursing journals appropriate for literature reviews. After reading this, the reader is clear on how to evaluate and conduct a literature review for nursing research and practice.

The book is organized into three parts. Part 1 discusses the importance of research. Part 2 focuses on planning a research study. Part 3 provides excellent strategies for evaluating, critiquing, and utilizing research in nursing practice; however, tips and questions to consider when critiquing research are woven throughout the book, appearing in each chapter, not just in Part 3. Each chapter opens with learning objectives and a glossary of key items, then closes with a summary of key ideas followed by learning activities.

This revised second edition includes two new chapters. The first of these, "Applying Appropriate Theories and Conceptual Models," is appropriately titled, for Fain shows the reader how to apply theory in research — a difficult connection to explain to nursing students. The discussion of theories is interesting and easy to understand, and puts theory into a practical perspective for nurses concerned more with practice than with the theoretical development of the discipline of nursing. The second new chapter focuses solely on evidence-based practice (EBP). After thor-

oughly discussing EBP for nursing practice, Fain provides tips on how to seek the best evidence for nursing practice.

Chapters 9 and 10, which cover data collection methods and data analysis, are the book's only weakness. They are incomplete, and therefore *Reading, Understanding, and Applying Nursing Research* should not be relied upon as the main textbook on these topics. This reader found misleading statements regarding some of the differences between quantitative and qualitative methods in Chapter 9. However, Chapters 12, 13, and 14, on phenomenological, ethnographic, and grounded theory research, are excellent. They are in-depth and extremely interesting, sure to captivate both students and instructors.

The problem with Chapter 10, "Analyzing the Data," is that multivariate tests of analyses, such as multiple and logistic regression, are not mentioned. These statistical tests are commonly used in clinically significant health-care research and should be included. Additionally, the term "odds ratio," a common one in health-care research, is not mentioned. Despite this particular shortcoming, the book is worthwhile. Instructors can provide supplementary materials.

In conclusion, this volume focuses on providing students with the skills they need to become educated and wise consumers of research in their clinical practice. Fain has provided an important resource for those of us who teach research to nursing students. It is a challenge to get students excited about research. Yet this book holds promise. It is geared to nurses who are focused on developing excellence in clinical practice. The breadth and depth of the research examples are substantive enough to pique students' interest in nursing research and to ensure that they understand and respect it. I highly recommend *Reading, Understanding, and Applying Nursing Research* as a required textbook for nursing research.

Joan Rosen Bloch, PhD, CRNP, is Assistant Professor, School of Nursing, College of New Jersey, Ewing, New Jersey, USA.