What’s the Rush?
Internal Validity Is Not a Trivial Pursuit

The goal of gerontological nursing research is to improve the health and well-being of older adults. Yet, a variety of critics have faulted this and other health care research for failing to influence public health policy and decision making. Research, critics say, has produced conflicting findings and esoteric reports that are not useful to public policy makers or clinicians. The conditions under which a gerontological nursing intervention is researched may be quite different than the conditions under which a policy is implemented; therefore, implementing an intervention program on a large scale may yield different outcomes than found in the original study. As one public official at a department on aging told me in frustration, “I want a phone number I can call where the researcher on the other end will give me the answers to the questions I have or will get me the answers.” In other words, there is an expectation that the research community will produce a solution or “bottom line” evidence that is useful to policy makers. This expectation is understandable given the complex health problems facing older adults that desperately need solutions. This editorial presents some thoughts on protecting the science of discovery in gerontological nursing, while being responsive to the need for generalizable interventions that can inform policy.

Federal agencies charged with health care research consider the nexus of research and public policy a priority. The Agency for Healthcare Research and Quality (AHRQ) (2002), for example, states in its mission: “AHRQ develops the evidence about what works in health care practice that enables providers, patients and consumers, system managers, purchasers, and policymakers to obtain real value for their health care dollar” (para. 1). The mission statement of the National Institutes of Health (NIH) also addresses the need to apply medical and behavioral knowledge generated through science to extend healthy life and reduce the burdens of illness and disability. The goals of NIH (2009b) include public accountability and social responsibility in the conduct of science. There is also an important time element that underscores the urgency of improving the connections between research and policy: The transfer of knowledge from research to practice takes an average of 15 years to achieve a use rate of 50% (Balas & Boren, 2000).

RANDOMIZED CONTROLLED TRIALS

To create evidence that is useful to public policy makers, research results must have applicability for larger groups than commonly seen in small, well-controlled randomized controlled trials (RCTs). Advances in technology and statistics provide greater opportunities to study larger, more representative samples. Computer technologies now provide access to large existing datasets of health care records. Sophisticated multivariate statistical modeling allows risk adjustments, control of sample selection biases, and examination of causal paths through the use of data collected in non-experimental studies. Should gerontological nurse researchers respond to the urgent need for generalizable solutions to health care problems by increasing the emphasis on generating knowledge through large-scale studies? Is the traditional four-phase clinical trial mechanism serving the needs of older adults, or should it be adapted or replaced?

Causal Modeling

First, let us examine the question of using causal modeling in non-experimental studies to generate causal knowledge. Causal modeling allows researchers to specify the whole model, mediating processes and short- and long-term outcomes associated with actions or interventions. With RCTs, there is often a decreased ability to examine long-term outcomes, and studying only proximal outcomes may provide an incomplete understanding of the usefulness or lack of
usefulness of a specific intervention. However, when using causal modeling, several different models can fit one dataset. The methods are prone to specification error and reciprocal feedback, and contingencies may dramatically change results. Large non-experimental studies often rely on survey data or existing datasets that are prone to error and superficiality. The validity of the findings that emerge from these data is suspect. Causal modeling has a purpose and is one more tool to advance understanding of the nature and behavior of complex and heterogeneous older adults. But causal modeling techniques applied to non-experimental data cannot adequately control multiple threats to internal validity, making the purpose of these methodologies different than in an RCT.

**Multiple Sites**

Next, let us examine conducting Phase II or III RCTs at multiple sites across larger regions so the intervention can be tested with a wider, more generalizable population. For highly standardized interventions with outcome measures that are sensitive to change, it is often feasible to use multiple sites because researchers can achieve procedural control and reliable measurement. For example, once safety has been established, new drugs and medical devices can be tested at multiple sites using highly standardized experimental protocols. However, it may be more difficult to achieve the degree of standardization and procedural control required in RCTs with many gerontological nursing interventions. Patient-centered nursing interventions usually have multiple assessment and treatment components. If multiple interventionists are used at different sites across the country, control of treatment fidelity becomes more difficult, and the intervention may be delivered differently, improperly, or with a bias for or against certain aspects of an intervention. The validity of the outcomes of intervention studies is based in large part on the degree to which researchers can maintain intervention fidelity. In addition, the potential for confounding influences from different organizational characteristics, regional mores, and historical factors across sites and regions increases the need for measurement and control of these factors.

It may also be difficult to achieve high-quality measurement of nursing outcomes across multiple sites. RCTs often require measurement over multiple times using tools that are reliable and highly sensitive to change. This kind of data collection is time intensive and costly. The ability to collect data on the same variables across the country is now made possible through the use of large, existing datasets (e.g., Outcome and Assessment Data Set, Minimum Data Set). Multiple health care workers have collected such records for purposes other than research. The superficiality, lack of sensitivity to change, and poor reliability of these measures make this kind of data poorly suited for experimental studies and increase the risks of Type I and Type II errors.

**Four Phases of RCTs**

The history of medical research provides many examples of the dangers of inadequate research for public health, including debacles such as thalidomide (Fine, 1972), bloodletting as a cure for multiple ailments (Warner, 1980), and thoracoplasty surgical treatment for tuberculosis (Tatak, 2004). It is within the four-phase clinical trial mechanisms that researchers protect older adults from the dangers of bad treatment.

The sagacity of the phases of RCTs lies in the validity, ethics, and integrity of focusing first on safety in early phase I trials, then on internal validity when establishing efficacy in Phases II and III, and finally on external validity during Phase IV to establish effectiveness in the general population. Trying to merge or rush through these phases is not good science. A well-designed RCT puts into service the basic experimental procedures taught to all beginning researchers: random assignment to groups; control procedures; the use of measures sensitive enough to capture change; and attention to construct, internal, and statistical-conclusion validity. Using these methods, claims of an intervention causing a specific outcome become more trustworthy. Scientists have a responsibility to use methods that will establish both safety and internal validity of interventions prior to any attempts to test interventions with larger populations. Only then can we protect both safety and the discovery of new knowledge, being assured an intervention has been tested using the most rigorous methodologies available.

**IMPLEMENTATION STUDIES**

In addition to protecting the essential four-phase sequence in intervention research, implementation studies are critical to achieve sustainable interventions that inform policy and improve the public’s health. Implementation research examines the use of strategies to adopt and integrate evidence-based health interventions and change practice patterns. As noted in a recent statement from the NIH (2009a), the approach to research dissemination in which interventions are transferred into a service setting without modification and based on a unidirectional flow of information from researchers to practitioners is insufficient to achieve practice change. Recent NIH funding opportunity announcements provide incentive and capacity for conducting implementation studies. The NIH is seeking to fund implementation research that moves away from a “top-down” approach to dissemination to interdisciplinary teams of scientists and practice stakeholders working together to develop and test conceptual models.
of dissemination and implementation that may be applicable across diverse practice settings.

CONCLUSION
While the title of this editorial states that internal validity is not a trivial pursuit, the same can be said for external validity. Research has often been described as the building of edifices one brick at a time, with each brick being a research study (Forscher, 1963). However, the ultimate goal is to build the edifice—that is, to achieve useful knowledge for large populations that can be translated into good public policies. If the bricks are weak, the edifice will crumble. Well-constructed edifices require strong bricks. Rigorously focusing on internal validity means gerontological nurse researchers must not rush to achieve external validity, so that ultimately, both the bricks and edifices will be sound.

REFERENCES

Christine R. Kovach, PhD, RN, FAAN
Professor
Methods Core Director, Self-Management Science Center
College of Nursing
University of Wisconsin-Milwaukee
Milwaukee, Wisconsin

The author discloses that she has no significant financial interests in any product or class of products discussed directly or indirectly in this activity, including research support.

doi:10.3928/19404921-20091214-01