A Few Thoughts on Measurement Issues

Every gerontological nurse investigator has a story about instrumentation. Whether we seek to describe physical or mental function associated with a specific chronic illness, untangle the effects of disease progression from age-related change, or track the effects of a nursing intervention, finding measures that adequately operationalize these constructs with older adults can be challenging. Some of these instrumentation issues stem from the rapidly changing face of the 65+ population in the United States, but some stem from the conceptual and methodological limitations of the instruments currently used. Each of these issues presents its own set of challenges.

THE CHANGING FACE OF THE OLDER ADULT POPULATION

Every day for the next 17 years, 10,000 adults in the United States will turn 65 (Pew Research Center, 2009). This older adult group will continue to be the fastest-growing population segment in the United States as the number of Baby Boomers who have reached age 65 continues to increase (Vincent & Velkoff, 2010). The increasing diversity of the older adult population segment (e.g., more minorities, a broader range of education and income, more varied family composition, management of comorbid conditions over a longer life span) are reflected in the aging populations of countries around the globe. Whether we are sufficiently prepared to capture these changing demographics in our target population depends to a great extent on the accuracy and dependability of existing measures.

LIMITATIONS OF EXISTING MEASURES

Because using new data collection measures limits comparing findings with those of existing studies, in which older, more established measures were used, the conventional advice we give to our neophyte investigators is to use data collection measures that have demonstrated adequate performance in prior studies (i.e., with reliability and validity coefficients within a defined range of acceptability). Although satisfactory answers to questions about psychometric performance characteristics (e.g., internal consistency and test-retest coefficients, factor loadings) of measures used in prior studies is valuable, an important first question to ask when choosing data collection measures is whether there continues to be “fit” between the measure and the population characteristics (Switzer, Wisniewski, Belle, Dew, & Schultz, 1999).

Health status is a construct that illustrates the difficulties of ensuring this fit. While health status is a variable in many gerontological nursing studies, there is no universally accepted definition of the construct, and investigators planning to include this construct in studies are faced with decisions about how best to operationalize it.

Domain-Specific Measures

Many investigators use instruments that measure functional capacity to operationalize health status (e.g., How many flights of stairs can you climb before you are short of breath? How far can you walk in 6 minutes?). Others choose measures that quantify the presence or absence of symptoms (e.g., fatigue, shortness of breath, pain). Some investigators choose measures consisting of items that tap individuals’ subjective feelings about their health status (e.g., I am as healthy as anyone else my age), whereas others focus on biometric indicators of health status (e.g., weight fluctuations, blood pressure changes, variations in body mass index, numbers and types of blood cells).

Domain-specific health status measures enhance comparisons of individuals/groups with similar health problems, such as older adults with specific cancers or chronic respiratory diseases like emphysema. However, the narrow focus of these measures limits comparisons across groups.

Global Measures

A second and increasingly more common approach to operationalizing a multidimensional construct like health status is to use a global measure—one that combines
subscales of items that assess various dimensions of the construct (e.g., scales assessing the older adult’s physical/mental health capacity and symptoms, as well as their subjective feelings about their own health). However, global measures that yield a single sum score make it difficult to determine the contributions of those separate dimensions in a multidimensional construct. In contrast, global measures composed of subscales with varying numbers of items and scoring metrics present a challenge in determining the predictable ability of individual subscales for explaining variations in health status. Further, the current approach to scoring global as well as domain-specific health status measure complicates score interpretation.

In that much of our research has been developed with the intent to improve nursing practice, many of the measures we use originated in the clinical practice setting. The scoring rubric for these “clinimetric” measures typically involves aggregation of item scores by summing or taking the mean of the item set. This scoring approach is based on the assumption that each item is equally important in explaining health status (Fayers & Hand, 2002). The questionable soundness of this assumption is easily demonstrated; for example, when an item assessing the occurrence of chest pain is weighted equally with an item assessing the occurrence of fatigue on a health status symptom measure, or when an insomnia item and a suicidal ideation item are given equal weight on a mental health status measure.

While health status measures have been used to illustrate the challenges faced in selecting from and using existing measures, the difficulties of operationalizing this construct are not unique. Each of us can think of other examples in which the plasticity of the constructs we use, and the limitations of existing measures for operationalizing those constructs with an older adult population, complicate achieving consensus within our research community.

MENTORING GERONTOLOGICAL NURSE INVESTIGATORS

We often fail to pass along what we learn about instrument issues to neophytes in our field. Although our investigators strive to increase the diversity of their study samples, it is often unclear from our reports how we adjusted our study designs to manage that diversity in our studies. While participant sex and race/ethnicity are well described in our study inclusion criteria and participant profiles, our study reports seldom mention whether and how these characteristics influenced our study measurement plans. For example, although cultural and educational differences are known to influence scores on cognitive function measures, our reports rarely include information on whether and how scoring rubrics for these measures were adjusted to reflect the cultural and educational diversity of our older adult samples.

At times, we are not explicit about who our study participants are. With the possible exception of the pediatric scientist, our investigators are the ones most likely to depend on data provided by proxy respondents—the older adult’s spouse, adult children, friends, and neighbors—when the adult is unable or unwilling to provide those data. Despite research findings indicating that proxy respondents’ reports of older adults’ physical and cognitive function and mood state are influenced by the proxy’s own physical and psychological health as well as the quality of their relationship with the older adult (Davis, 2001), information on the characteristics and credibility of these proxy respondents is often omitted from our reports.

DEVELOPING THE MEASUREMENT AGENDA

Given the rapidly changing face of the older adult population in the United States and the limitations of existing measures, there is sufficient justification to revisit the measurement science underpinning gerontological nursing. Now is the time to formulate a gerontological measurement agenda to redefine the constructs we study, provide opportunities for exploration of and thoughtful discussion on the relative strengths and limitations of existing measures, and provide direction for where we should put our energy and resources in developing and testing future gerontological research measures. For example, we might consider:

- Revisiting existing data on the measures we use in our studies, including information on their origins, and whether these instruments continue to fit the characteristics of the current older adult population.
- Including information in our published reports on how and from whom data on older adult study participants are collected.
- Publishing research briefs in our research journals related to when and how we adjust measure scoring rubrics for sex, educational, and cultural differences in our samples, as well as the outcomes of those adjustments.
- Convening periodic consensus conferences to establish priority areas for measure development.
- Submitting measurement papers for national and regional gerontology conferences with the intent of accelerating the dissemination of recent findings from measurement studies.
• Developing a national repository of measures with a history of successful testing with gender, culturally, and educationally diverse older adult groups.

Whether we are sufficiently prepared to capture the changing demographics of the older adult population in the United States depends on both the accuracy and dependability of the measures we use, as well as our commitment to mentor the future generation of gerontological nurse investigators about the use of those measures.

REFERENCES

Linda Lindsey Davis, PhD, RN, ANP, FAAN
Gardiner Professor, School of Nursing
Senior Fellow
Center for the Study of Aging and Human Development
Duke University
Durham, North Carolina

The author has disclosed no potential conflicts of interest, financial or otherwise.

doi:10.3928/19404921-20131017-01