

DOUBLE ISSUE

The **BEHAVIORAL MEASUREMENT** *Letter*

**Behavioral
Measurement
Database
Services**

Enriching the health and behavioral sciences by broadening instrument access

Vol. 7, No. 1
Winter 2000

Introduction to This Issue

This issue of *The Behavioral Measurement Letter* contains a very large volume of material, so large, in fact, that it is a double issue. And although I was again assisted by our regular contributor, Fred Bryant, there was a related delay in its publication. I hope that both the quantity, and especially the quality of its contents will be adequate compensation for the wait.

In this issue of The BML, Dr. Frank Baker, Director of Research at the American Cancer Society, reviews instruments of various types used to measure the quality of life (QOL) of cancer survivors and research designs for evaluating cancer survivors' QOL, and then discusses challenges and issues in QOL research. Due to the implementation of cancer screening programs and lifestyle changes (e.g., in diet, cigarette smoking, alcohol consumption), improved diagnostic methods, and the use of new and more effective cancer treatments in the last quarter of the 20th century, cancer, still associated with high levels of morbidity and mortality, is no longer the dreaded killer it once was. These advances have resulted in large and growing numbers of cancer survivors, many of whom live for years (even decades) beyond the time of initial diagnosis. Thus, instruments and methods are needed to assess the quality of life of cancer survivors in order to obtain baseline measures, to identify factors that contribute to a good QOL and those that contribute to a poor QOL for survivors, to design and improve ways and means for QOL enhancement, and to determine the effectiveness of attempts to improve cancer survivors' QOL.

Racial, ethnic, and class bias are common, often substantial sources of error variance in measurement. Such bias may be introduced at any stage of instrument development and use, including definition and operationalization of variables to be measured, item construction, and instrument administration. These sources of measurement bias are discussed in a column by Drs. Mildred Ramirez, Marvella Ford, and Anita L. Stewart, from the Research Centers for Minority Aging Research -- Measurement and Methods Cores. They point out that measures administered to various racial/ethnic groups and/or persons of low socio-economic status that do not account for racial/ethnic/class differences can produce results that are not generalizable to these groups. This, in turn, leads to flawed social policies and ineffective services designed using such research. The column strongly reminds us that 1) measurement bias of various types exists and must be addressed effectively to assure validity, and 2) the type of measurement bias due to insensitivity to racial, ethnic, and/or class differences has consequences not only for the corpus of research-based knowledge, but for applications of such invalid knowledge.

Also in this issue is the second of a two-part column, "Culture . . . One Step at a Time," by cultural anthropologist John Gatewood. As the reader may recall, the first part (The BML (6) 2:5-10, Fall 1999) dealt with means to gain an understanding of a culture by learning how persons within the culture view their world. He presented three methods to discover how persons see similarities, differences, and relationships among products of a culture or items in its environment -- free listing, pile sorting, and triadic comparison. In the second

part of his column, Dr. Gatewood explores various ways to analyze data generated by these methods. Each is a type of consensus analysis and allows for analyses of either 1) similarities, differences, and relationships among items familiar to a culture, or 2) the level of knowledge each informant has about these items. (The reader is referred to part one for detailed discussion of the three methods and for references to software used to analyze data generated by these methods.)

Address comments and suggestions to The Editor, *The Behavioral Measurement Letter*, Behavioral Measurement Database Services, PO Box 110287, Pittsburgh, PA 15232-0787. If warranted and as space permits, your communication may appear as a letter to the editor. Whether published or not, your feedback will be attended to and appreciated.

We also accept short manuscripts for The BML. Submit, at any time, a brief article, opinion piece or book review on a BML-relevant topic to The Editor at the above address. Each submission will be given careful consideration for possible publication.

HaPI reading . . .

Al K. DeRoy, Editor

Staff

Director Evelyn Perloff, PhD
Newsletter Editor Al K. DeRoy, PhD
Record Indexer Jacqueline A. Reynolds, BA
Analyst Amanda Winner, MA
Computer Consultant Alfred A. Cecchetti, MS
Measurement Consultant Fred B. Bryant, PhD
Information Specialist
Consultant Ellen Detlefson, PhD
Copy Reviewer Tracey Handlovic, BA
Business Services Manager Diane Cadwell
File Processor Betty Hubbard
Phone: (412) 687-6850 Fax: (412) 687-5213
E-mail: bmdshapi@aol.com

There is nothing so unthinkable as thought,
unless it be the entire absence of thought.

Samuel Butler

Assessing the Quality of Life of Cancer Survivors

Frank Baker

As a result of improvements in prevention, early detection, and treatment, more people are surviving cancer today than at any time in the past. It is estimated that there are at least 8.4 million people in the United States who are alive today after receiving a cancer diagnosis (American Cancer Society, 2000). The image of cancer has changed from a disease that was viewed as synonymous with death, to that of a chronic illness that can be survived but can inherently change various aspects of the life of the person who has the disease.

Even the definition of who is considered a cancer survivor has changed in recent years. The traditional definition of a cancer survivor was someone who was alive five years after diagnosis and was disease-free (Friedman, 1980). This definition of five-year survival was used because for many types of cancer, if a patient survives for five years, it is likely that he or she is cured. With the development of the Office of Survivorship at the National Cancer Institute, a new definition was offered that defined cancer survivors as people who were living two years postdiagnosis (Meadows et al., 1998). The American Cancer Society (ACS) has offered a definition of survivorship that is more inclusive, i.e., cancer survival begins with diagnosis and continues for the balance of the person's life (H. Eyre, personal communication, April 24, 2000).

Cancer survivors have been described as going through a series of stages of survival (Mullan, 1985). First is an acute stage that begins at diagnosis and continues through the initial courses of medical treatment. At this stage, the survivor is commonly called a patient, and the primary focus is on physical survival. This period is followed by an extended, or intermediate, stage of survival in which survivors monitor their bodies for recurrence of disease. In this stage of recovery, quality of life becomes a major concern as survivors deal with the physical and emotional effects of treatment and their reentry into social roles. Most of the

behavioral and psychosocial research with adult survivors has focused on the initial stage, when patients are in active treatment, and to some lesser extent on the second stage, when patients are likely to be seen at university-related hospitals and cancer centers where there is grant support for this kind of research.

The third major stage is called sustained remission, or long-term survival. Some long-term cancer survivors will do well. Others, however, must deal with chronic and late effects of treatment that is often toxic, that produces debilitating symptoms and impaired functioning, and that leads to short- and long-term changes in an individual's ability to conduct a "normal" life. Thus, ever-improving technology and new treatments are saving cancer patients' lives but at the same time have side effects and toxicities that can radically affect the quality of their lives.

Traditionally, medical evaluation of cancer treatment outcomes included tumor response, disease-free survival, and overall survival (US Department of Health and Human Services/Public Health Service/National Institutes of Health, 1990). As length of survival increased, however, the value of cancer treatments began to be judged not only by which treatments contributed to increased survival, but also to the quality of survivors' lives. This new criterion was called "quality of life" (QOL), a term that received its own heading in Index Medicus in 1977 (Frank-Stromborg, 1984). Quality-of-life research has increased dramatically in recent years, and there are numerous examples of activities on the national and international levels that indicate growing recognition of the importance of including QOL in assessments of patient outcomes. For example, the United States Food and Drug Administration uses QOL measurements in the process of approving drugs for the treatment of cancer (Johnson & Temple, 1985), and the National Cancer Institute includes assessments of QOL in endpoint evaluations of cancer clinical trials (Moinpour et al., 1989). The American Cancer Society (ACS) has set national challenge goals for 2015 that include improving the quality of life of cancer

survivors in addition to decreasing cancer incidence and mortality (Beyers et al., 1999).

Early attempts to assess a cancer patient's quality of life were based on ratings by doctors or other caregivers of the patient's ability to function without special assistance. One of the first of these was the Karnofsky Performance Status Scale, which provides a rating of cancer patients' physical functioning on a scale from 0% to 100% (moribund - normal) in increments of 10 (Karnofsky & Burchenal, 1949). Although this global clinician rating has been used widely by cooperative groups involved in clinical trials because it shows correlations with tumor response and survival, studies have shown that it has low interrater reliability and that such clinician-based ratings correlate poorly with patients' ratings (Aaronson, 1990).

Over the last several decades, the view that the patient should be the primary source of information regarding his or her quality of life has gained general acceptance (Aaronson, Beckmann, Bernheim, & Zittoun, 1987; Cella & Tulsky, 1990). The trend toward using cancer patient self-report was furthered by studies that demonstrated that proxy ratings, whether by health care providers or family members, showed little agreement with patients' ratings and with each other as well (Epstein, Hall, Tognetti, Son, & Conant, 1989; Osoba, 1994). When physicians are used as the source of quality-of-life ratings, they tend to emphasize physiological data and focus on symptoms and physical functioning, whereas nurses, social workers, and family members tend to emphasize psychosocial dimensions (Schipper, 1990). However, no matter who is used as a proxy for the patient, they tend to produce ratings that underestimate the QOL of patients as rated by the patients themselves (Sprangers & Aaronson, 1992). Indeed, it has been suggested that one of the most important positive effects of the widespread acceptance of the quality-of-life concept in medical practice has been the recognition that the perspective of the patient is as valid as that of the clinician (Leplege & Hunt, 1997).

While no universally accepted definition for QOL has developed over time, a general consensus has grown that quality of life is best

defined as a complex multidimensional construct, not just in terms of a single dimension such as level of physical functioning (Donovan, Sanson-Fisher, & Redman, 1989; Grant et al., 1992). The World Health Organization defines quality of life as "individuals' perceptions of their position in the context of the culture and value system in which they live and in relation to their goals, standards, and concerns" (World Health Organization—Division of Mental Health, 1993). Six broad domains are included in their definition: physical health, psychological state, levels of independence, social relationships, environmental features, and spiritual concerns. Researchers focusing on cancer survivors have developed some consensus about a minimum of four dimensions to include in quality-of-life measures: physical functioning, emotional and psychological functioning, social functioning, and disease/treatment-related symptoms (Aaronson, 1990; Moinpour, Hayden, Thompson, Feigl, & Metch, 1990; Ganz, 1994). In addition, a global measure of perceived health status may enhance QOL measurement (Haberman & Bush, 1998).

Two approaches have been employed in attempts to define operationally the multiple dimensions of quality of life. One approach focuses on developing a single multidimensional scale, while a second approach is to develop many separate scales to measure QOL. The second approach may involve using a battery of single-domain scales to assess QOL, or a combination of a multidimensional measure with one or more scales that measure other domains. Instruments developed using the first approach often include generic mental health measures of such variables as depression, anxiety, mood state, self-esteem, perceived meaningfulness of life, and the like. In most cases, these measures have been normed with populations other than cancer patients, such as college students and outpatient psychiatric patients, and thus little or no information is available on how well they measure QOL of cancer patients. To overcome these limitations, normative data, especially data on validity and reliability, have been developed for some measures based on their administration to groups of cancer patients.

There have also been attempts to deal with these limitations by developing cancer-specific and diagnosis-specific QOL measures. While there is little theoretical basis for deciding what to include in a cancer-specific measure of QOL, most include measures of the key dimensions of physical well-being and associated symptoms, of psychological well-being, and measures of social well-being (Cella & Tulskey, 1993). Some researchers, too, argue for the inclusion of spirituality as a core dimension of QOL, pointing out that in interviews concerning quality of life, cancer survivors often emphasize the role of spirituality (Ferrell, Grant, Padilla, & Vemuri, 1991). In this regard, specific tools for measuring the spiritual well being of cancer patients are being developed, and evidence of a strong relationship of spirituality to QOL has been reported recently (Brady, Peterman, Fitchett, Mo, & Cella, 1999).

As interest in the assessment of quality-of-life treatment outcomes for cancer and other chronic diseases grew, some existing generic measures of health status outcomes were used initially. They were called measures of "health-related" quality of life (HRQOL), even though they had not originally been developed as QOL measures. These generic HRQOL measures include: the Sickness Impact Profile (SIP; Bergner, Bobbitt, Carter, & Gilson, 1981), which covers a wide range of functioning and focuses on behaviors rather than subjective expressions; the Medical Outcomes Study Short Form-36 (MOS SF-36; Stewart, Hays, & Ware, 1988), which consistently distinguishes among groups of patients differing in severity of illness; and the Nottingham Health Profile (NHP; Hunt, McKenna, McEwen, Williams, & Papp, 1981), which has most widely been used in the UK. It should be noted here that although the NHP is less precise in detecting distress associated with mild illness than the other generic HRQOL measures, it seems to assess adequately differences between groups with moderately severe to severe illness.

While generic measures are commonly used to assess outcomes in other chronic diseases, there has been significant focus on developing disease-specific measures for cancer outcomes.

For example, quality-of-life research on the effects of cardiovascular disease has generally relied on generic measures, with some dimension-specific measures included as part of a "battery-of-measures approach" in which multiple distinct instruments are used, each measuring a specific QOL domain and each being scored separately (National Institutes of Health/National Heart Lung & Blood Institute, 1995). Additionally, since cancer is actually a number of diseases affecting different sites in the body, there has been a further refinement of cancer QOL measures to develop site-specific scales as well.

Among the measures of quality of life developed specifically to use with cancer patients that have been shown to have adequate reliability and validity are the Cancer Rehabilitation Evaluation System (CARES; Schag, Heinrich, Aadland, & Ganz, 1990), the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC-QLQ-C30, Aaronson et al., 1993), the Functional Assessment of Cancer Therapy (FACT; Cella, et al., 1993), the Functional Living Index-Cancer (FLIC; Schipper, Clinch, McMurray, & Levitt, 1984), the Satisfaction with Life Domains Scale for Cancer (SLDS-C; Baker, Curbow, & Wingard, 1992), and the Spitzer Quality of Life Index (SQOL-Index; Spitzer et al., 1981). The original CARES included 139 problems that might be encountered by cancer patients on a daily basis as they attempted to cope with their disease and its treatment. A shorter form consisting of 59 items has been developed, the Cancer Rehabilitation Evaluation System—Short Form (CARES-SF; Schag, Ganz, & Heinrich, 1991). Unlike some of the other QOL measures for cancer, both forms of the CARES require permission for use and payment of fees, which may have limited their use. The European Organization for Research and Treatment of Cancer (EORTC), which involves 15 countries in one of the oldest and largest cancer clinical trials groups in Europe, developed the QLQ-C30 as a core instrument to which diagnosis-specific questionnaire modules could be added. The EORTC QLQ-C30 was developed specifically for assessment of the QOL of cancer patients

participating in multinational clinical trials and has numerous parallel language versions. The FLIC and the Spitzer SQOL-Index currently appear not to be as popular as when they were first developed in the 1980s. The SLDS-C, consisting of only 17 items and employing a "smiley-frowney" faces response format to assess different life domains, has been shown to be a relatively less burdensome cancer QOL measure for patients to complete (Baker, Jodrey, Zabora, Douglas, & Fernandez-Kelly, 1996). The FACT-G, a general version of The Functional Assessment of Cancer Therapy scale that can be used with patients of any tumor type, has gone through four revisions (Cella, 1997). Disease-specific versions of the FACT have been developed for many of the major types of cancer (i.e., breast, bladder, colorectal, head and neck, lung, ovarian, and prostate), as well as for HIV infection, by adding items specifically relevant to different cancer sites to the core set of items in the FACT-G (Kornblith & Holland, 1994).

Several measures have been developed to evaluate the unique quality-of-life outcomes of bone marrow transplantation, a particularly demanding cancer treatment. Among these are the Quality of Life Scale—Bone Marrow Transplant, developed at the City of Hope by Ferrell and Grant (Ferrell et al., 1992; Grant et al., 1992), the Cella FACT-BMT (McQuellon et al., 1997), and the Bush Bone Marrow Transplant Symptom Inventory (Bush, Haberman, & Donaldson, 1995; Bush & Langer, 1998).

Research Designs for Evaluating Quality of Life

The most commonly used research designs in psychosocial and behavioral research on cancer have been cross-sectional in nature. Cross-sectional assessments are useful for describing the current effects that survivors are experiencing at a particular point in time in relation to other variables assessed at that time. However, this type of design has several noteworthy deficiencies. Retrospective reports are of only limited reliability regarding the respondents' perceived quality of life at an earlier time, such as at the time of treatment or before cancer diagnosis. Retrospective

assessments performed years after treatment also have a selection bias in that only those who survived to that time are available to be asked about their experience of the disease and its treatment, while there is no comparable data on those who died.

Clinical trials involve random assignment into groups exposed to an intervention and groups not exposed, and the collection of data before and after exposure to the intervention. With the heightened attention being given to QOL effects of new treatments, QOL measures that are sensitive to change have become of particular interest to pharmaceutical companies for use in clinical trials of drug effectiveness. Unfortunately, these trials may not show the positive QOL effects that are hoped for because of lack of attention to, or understanding of, the psychometric characteristics of particular measures or what dimensions they include. Another limitation to the research designs that have been used traditionally to evaluate quality of life of cancer patients is their predominant dependence on relatively small opportunistic samples of patients from large cancer treatment centers. These samples have usually had little diversity, tending to exclude racial/ethnic minority group members, individuals from rural areas, and people of lower socio-economic status.

The use of prospective, population-based, longitudinal designs with representative samples of people diagnosed with different types of cancer selected by using state cancer registries as a sampling frame offers a way of overcoming a number of these deficiencies.

Cross-Cultural Health-Related Quality-of-Life Instruments

Over the last decade, interest has grown significantly in developing measures that can be used to compare quality of life in different countries, cultures, and languages (Anderson, Aaronson, Bullinger, & McBee, 1996). This has been further stimulated by the trend toward multinational pharmaceutical trials that require measuring the same constructs cross-culturally

with the same degree of precision to permit pooling data from different countries for analysis.

In the United States, as clinical trials funded by the National Institutes of Health required the inclusion of minorities wherever possible, it was found that quality-of-life measures that had been developed using White middle-class populations appeared to be difficult for poor and minority populations to respond to and that these measures did not seem to be culturally relevant for some groups (Baker et al., 1996). The National Cancer Institute therefore funded a number of research projects to develop measures that were appropriate for minorities and lower socioeconomic groups (e.g., Warnecke et al., 1996). In a study that sought to test the cultural equivalence of QOL instruments that might be used with Hispanic and African American cancer patients, it was found that the measures that seemed most appropriate for these groups used simple, often graphic, response formats (Baker et al., 1996). For Hispanic populations, the most successful technique involved translation into Spanish and back-translation into English by individuals who spoke the varieties of Spanish associated with subjects' different countries of origin. Some quality-of-life measures turned out not to be culturally relevant or acceptable to some minority groups. For those with limited literacy, it was found that a brief pretest to determine level of literacy was useful, and that administering the QOL measures through an interview was desirable.

Quality-of-Life Measures for Children

Unlike adult cancer survivors who have been studied for the most part with cross-sectional research designs, child survivors have been followed in long-term longitudinal studies. Until recently, however, quality-of-life research with pediatric cancer survivors has been limited by the fact that most quality-of-life measures were designed for adults and were not appropriate for children. Thus, instead, parents or health care providers were asked to report on children's quality of life (Bradlyn et al., 1996). As noted above, however, proxy assessments of a patient's quality of life rarely agree with the patient's self-report. Thus, in recent years, a serious commitment has been made to

developing measures that are based on qualitative explorations of the pediatric perspective and that are developmentally appropriate. As a result, creative formats for self-report measures are being called for in assessing the QOL of children and adolescents (Hinds & Haase, 1998). Although visual analog scales and other graphic formats, for example, have increased the choice of methods for presenting questions, developing games and interactive videos to ask questions about QOL might better maintain children's motivation and willingness to respond to the same items repeatedly over months or years.

Some of the same problems in quality-of-life assessment of persons with low literacy are also encountered with children, and some of the same solutions are appropriate for use with children. As noted above, until just a few years ago, assessment of quality of life for childhood cancer patients was dependent on the responses of their parents or caregivers. Proxy respondents were used even for older children able to read and understand items developed for adults. In recent years, these "adult-normed" measures are adding normative data developed for teen-age and young adults populations. Moreover, as mentioned above, innovative techniques using presentation formats appropriate for young children are being developed as well.

Respondent Burden

A constraint on what and how many QOL instruments to use is the issue of respondent burden. Cancer patients must deal with significant fatigue, nausea, pain, and other symptoms that make it difficult for them to complete long, burdensome questionnaires and/or interviews. But reliability and sensitivity to change are usually enhanced as instrument length increases. Also, covering multiple dimensions of quality of life usually requires use of multiple scales or measures, either in one instrument or as a battery. In addition, in order to track changes in quality of life over time, a measure must be repeated over time. Thus, in assessing the QOL of cancer patients, researchers and clinicians often must struggle to gain a reasonable balance between maximizing

coverage, reliability and sensitivity on the one hand, and minimizing respondent burden on the other.

Furthermore, respondents may become frustrated with having to answer the same or similar questions repeatedly. Ordering the presentation of questions logically seems helpful, as does labeling each section of the instrument. Grouping scales that have similar response formats may also be helpful, since patient respondents may become irritated with shifting among response formats employed in different scales. Patients also complain about the changing time periods that different scales use -- a week, two weeks, a month, six months, a year, and so forth, as well as shifting back and forth across formats. Pilot-testing a questionnaire or interview with a few representative patients is critical to resolving these issues.

Asking a cancer patient to respond to questions about various aspects of their life is not, however, all burden. Often patients say that they appreciate the researcher's interest in their personal experience and in having an opportunity to describe any QOL problems that they may be facing. Some patients have described the experience of responding to QOL questions as having a therapeutic effect that helps them place their experiences in perspective and broadens their consideration of what is happening to them and their families. Responding to questions about their lives seems to have healing effects, which have been described in relation to the "healing power" of telling one's story, by providing cognitive, affective, interpersonal, and personal growth benefits (Heiney, 1995).

Some relatively short instruments only take two or three minutes to complete. One of these is Cantril's Self-Anchoring Ladder of Life [Self-Striving Scale], which was designed to measure a general sense of well-being (Cantril, 1965). This measure, which uses a simple graphic response format, has high content validity as well as face validity. Respondents are presented with a picture of a 10-rung ladder with the highest rung marked "10" and the lowest rung marked "1". They are then asked to indicate where they would place themselves now, some specific time in the past, and some time in the

future in terms of their satisfaction with their lives. This approach has been used successfully in assessing well-being with cancer patients undergoing bone marrow transplantation (Baker, Curbow, & Wingard, 1991; Baker, Zabora, Jodrey, Polland, & Marcellus, 1995) and with patients receiving other treatments (Ashing-Giwa & Ganz, 1997; Visser & Smets, 1998).

Quality-of-Life Surveillance

There has been growing recognition by the public health community that a focus on biomedically-oriented measures of a population's well-being, such as mortality and morbidity, provides an inadequate picture of the outcomes of prevention and treatment, and that additional measures are needed. For example, both the earlier "Healthy People 2000" and the updated goals for the nation in "Healthy People 2010" include not only goals to prevent premature death, disability, and disease, but also goals concerning enhancement of the quality of life. Population-focused goals such as these pose unique challenges for QOL assessments, which have tended to be limited to collecting data from opportunistic samples of people.

Increasingly progress is being made in including QOL measures in large population-based studies. In 1993, measures of health-related quality of life (HRQOL) were introduced into the Centers for Disease Control and Prevention's Behavioral Risk Factor Surveillance System (BRFSS), which uses telephone surveys to monitor health risk behaviors among adults (Centers for Disease Control and Prevention, 1995). Four general questions are included: (a) ratings of the respondent's health from "excellent" to "poor"; (b) how many days during the previous 30 days the respondent's physical health was not good because of injury or illness; (c) how many days during the previous 30 days their mental health was not good because of stress, depression, or "problems with emotions"; and (d) how many days during the previous 30 days their physical or mental health prevented them from performing usual activities, such as self-care, work, or recreation. Since 1995, state health departments have had the option of

including an expanded set of 14 HRQOL questions in the BRFSS.

As noted above, the ACS, which has set goals for improving the quality of life of cancer survivors by the year 2015, is developing a nation-wide longitudinal, prospective study that is seeking to enroll a population-based sample of tens of thousands of cancer survivors to determine their changing quality of life and needs for services (Denniston, Brogan, Baker, Dye, & Hann, 1999). In this ACS study, a battery of quality-of-life measures is being employed in questionnaire surveys, repeated over a ten-year period, of people who survive one of the ten most common cancers.

The ACS is also conducting secondary analyses of data available from existing national surveys to provide a baseline assessment of the current quality of life of cancer survivors and to determine how the ACS can best allocate its resources to meet the goal of improving the QOL of cancer survivors. Data from the 1998 Health Care Financing Administration's (HCFA's) Medicare Health Outcomes Survey (MHOS) of people enrolled in managed care services through Medicare were analyzed to provide baseline estimates of quality of life of cancer survivors in comparison to a matched cohort of respondents who had not suffered cancer. The HOS was mailed to a random sample of 1,000 Medicare beneficiaries continuously enrolled for six months in a plan with a Medicare contract in place on or before January 1, 1997. In plans with 1,000 or fewer Medicare enrollees, all eligible members were surveyed. The 1998 data collection included 279,135 beneficiaries enrolled in 268 managed care plans covering 287 market areas. Completed surveys were received from 167,096 respondents for a raw response rate of 60%. An analytic data set was created that merged together plan characteristics from HCFA, administrative data files containing the survey response data, and patient demographic and entitlement information from the Medicare Enrollment Data Base. Some 23,137 respondents who answered positively when asked if a doctor had ever told them that they had any cancer other than skin cancer were age-matched to an equal number of respondents who indicated that they had not been diagnosed with

cancer. Comparisons of the cancer survivors with the noncancer group indicated that the cancer survivors had significantly poorer scores on all eight of the subscales of the Medical Outcomes Study Short Form-36 (MOS SF-36; Baker, Hafter, & Denniston, 2000). Other analyses of cancer survivors' quality of life by ACS in collaboration with HCFA are planned for the continuing HCFA surveys of the quality-of-life outcome in Medicare-contracted managed care.

QOL Issues for Future Research

Scales used for assessing the quality of life of cancer patients often suffer from ceiling and floor effects. If there is no room for respondents to indicate improvement for their quality of life because respondents tend to cluster at the top scores on the scale (ceiling effect), the measure will have limited utility in assessing positive changes in QOL over repeated assessments. Similarly if the scale does not allow a respondent to indicate a decrease in quality of life because respondents tend to be clustered at the bottom of the scale with little room for downward change (floor effect), there is a question of how useful the scale can be. Both problems have been encountered in using various quality-of-life measures to assess the QOL of cancer patients. For example, the MOS SF-36 has been observed to show some floor effects in that it does not sufficiently allow the very old to indicate limits in their physical functioning. Given that cancer tends to come later in life (American Cancer Society, 2000), this poses a particular problem in assessing the QOL of longer-term survivors who may be of advanced age.

Related to ceiling effects is the common finding that cancer survivors rate their quality of life at very high levels. Some studies that have compared the responses on self-report measures of quality of life of cancer patients and normal, healthy population groups have found that the cancer patients fail to show a lower quality of life than the healthy individuals (Breetvelt & van Dam, 1991), as opposed to results from more in-depth interviews with these patients and

the experience of physicians, nurses, and other caregivers with these patients. This apparent over-estimation in patient self-reporting has also been found with other patient groups. The phenomenon, termed "response-shift" by some researchers, is attributed not so much to limitations of a specific measure, but rather to a change in standard used by survivors, i.e., they are so happy to be alive that they shift the bases for their comparisons in responding to quality-of-life questions.

Psychosocial and behavioral research, including QOL research, have tended to concentrate on breast cancer for various reasons, including the facts that this type of cancer is the most common cancer in women, that it threatens an organ that is associated with psychological issues of prime importance to both sick and healthy women, and that there generally has been greater availability of funding for research on this type of cancer than others. Recently, more attention has been given to prostate and colorectal cancer, two of the other most commonly occurring types of cancer, and consequently more attention is being given to the unique QOL issues associated with these cancers and other types as well.

Given that there is no "gold standard" for quality-of-life assessment, there is a tendency toward using *ad hoc* measures, so it is difficult to compare results across various studies. Thus, the inclusion of QOL measures as a part of population-based surveillance of the general population is an asset to be further explored and developed. Also, increasing the number of cancer survivors in such population surveys will allow tracking of improvements associated with national efforts to improve QOL by various organizations over the next decades.

Moreover, greater efforts to develop consensus regarding the conceptualization and measurement of QOL in cancer certainly seem to be justified. These efforts should involve more participation of cancer patients and survivors in defining their QOL. Further development of conceptual and theoretical models of QOL in cancer is needed to guide research in this area. The absence of conceptual development is particularly evident in the assessment of quality of life of childhood cancer patients, survivor populations, and family members of cancer patients/survivors.

Assessing the Quality of Life of Cancer Survivors

(continued)

Finally, quality of life at the end of life is gaining increasing attention. This type of assessment poses unique methodological problems that require special attention. For example, given that one of the major problems of dying cancer patients is severe pain, medication to manage the pain may make it difficult or impossible to collect self-report QOL data. In these cases it may be necessary to use proxy respondents, which poses the problems noted above. It also may be necessary to identify different life concerns at the end of life, such as the patient's needs to feel that he or she is being treated with dignity, has settled their affairs, and has resolved strained interpersonal relationships (Stewart, Teno, Patrick, & Lynn, 1999).

References

Aaronson, N.K. (1990). Quality of life research in cancer clinical trials: A need for common rules and language. *Oncology, 4*, 59-66.

Aaronson, N.K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N.J., Filiberti, A., Flechtner, H., Fleishman, S.B., deHaes, J.C.J.M., Kaasa, S., Klee, M., Osoba, D., Razavi, D., Rofe, P.B., Schraub, S., Sneeuw, K., Sullivan, M., & Takeda, F. (1993). European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC-QLQ30). *Journal of the National Cancer Institute, 85*, 365-376.

Aaronson, N.K., Beckmann, J.H., Bernheim, J.L., & Zittoun, R. (Eds.). (1987). *The quality of life of cancer patients*. New York: Raven Press.

American Cancer Society. (2000). *Cancer facts and figures, 2000*. Atlanta, GA: Author.

Anderson, R.T., Aaronson, N.K., Bullinger, M., & McBee, W.L. (1996). A review of the progress towards developing health-related quality-of-life instruments for international clinical studies and outcomes research. *PharmacoEconomics, 10*, 336-355.

Ashing-Giwa, K., & Ganz, P.A. (1997). Understanding the breast cancer experience of African-American women. *Journal of Psychosocial Oncology, 15*, 19-35.

Baker, F., Curbow, B., & Wingard, J.R. (1991). Role retention and quality of life of bone marrow transplant survivors. *Social Science & Medicine, 6*, 697-704.

Baker, F., Curbow, B., & Wingard, J.R. (1992). Development of the Satisfaction with Life Domains Scale

for Cancer. *Journal of Psychosocial Oncology, 10*(3), 75-90.

Baker, F., Haffer, C., & Denniston, M. (2000, January). *The health-related quality of life of cancer survivors in Medicare managed care*. Washington, DC: Paper presented at the Partnerships for Health in the New Millennium: Launching Healthy People 2010 Conference.

Baker, F., Jodrey, D., Zabora, J., Douglas, C., & Fernandez-Kelly, P. (1996). Empirically selected instruments for measuring quality-of-life dimensions in culturally diverse populations. *Journal of the National Cancer Institute, 20*, 39-47.

Baker, F., Zabora, J., Jodrey, D., Polland, A., & Marcellus, D. (1995). Quality of life and social support of patients being evaluated for bone marrow transplantation. *Journal of Clinical Psychology in Medical Settings, 2*, 357-372.

Bergner, M., Bobbitt, R.A., Carter, W.B., & Gilson, B.S. (1981). The Sickness Impact Profile: Development and final revision of a health status measure. *Medical Care, 19*, 787-806.

Byers, T., Mouchawar, J., Marks, J., Cady, B., Lins, N., Swanson, G.M., Bal, D.G., & Eyre, H. (1999). The American Cancer Society challenge goals. How far can cancer rates decline in the US by the year 2015? *Cancer, 86*, 715-727.

Bradlyn, A.S., Ritchey, A.K., Harris, C.V., Moore, I.M., O'Brien, R.T., Parsons, S.K., Patterson, K., & Pollock, B.H. (1996). Quality of life research in pediatric oncology. Research methods and barriers. *Cancer, 78*, 1333-1339.

Brady, M.J., Peterman, A.H., Fitchett, G., Mo, M., & Cella, D. (1999). A case for including spirituality in quality of life measurement in oncology. *Psycho-Oncology, 8*, 417-428.

Breetvelt, I.S., & van Dam, F.S.A.M. (1991). Underreporting by cancer patients: The case of response-shift. *Social Science & Medicine, 32*, 981-987.

Bush, N.E., Haberman, M., & Donaldson, G. (1995). Quality of life of 125 adults surviving 6-18 years after bone marrow transplantation. *Social Science & Medicine, 40*, 479-490.

Bush, R.A., & Langer, R.D. (1998). The effects of insurance coverage and ethnicity on mammography utilization in a postmenopausal population. *Western Journal of Medicine, 168*, 236-240.

Cantril, H. (1965). *The pattern of human concerns*. New Brunswick, NJ: Rutgers University Press.

Cella, D.F. (1997). *The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System*. Evanston, IL: Center on Outcomes Research & Education (CORE), Evanston Northwestern Health Care, Northwestern University.

Assessing the Quality of Life of Cancer Survivors

(continued)

- Cella, D.F., & Tulsky, D.S. (1990). Measuring quality of life today: Methodological aspects. *Oncology, 4*, 29-38.
- Cella, D.F., & Tulsky, D.S. (1993). Quality of life in cancer: Definition, purpose, and method of measurement. *Cancer Investigation, 11*, 327-336.
- Cella, D.F., Tulsky, D.S., Gray, G., Sarafian, B., Linn, E., Bonomi, A., Silberman, M., Yellen, S.B., Winicour, P., Brannon, J., & et al. (1993). The Functional Assessment of Cancer Therapy Scale. Development and validation of the general measure. *Journal of Clinical Oncology, 11*, 570-579.
- Centers for Disease Control and Prevention. (1995). Health-related quality-of-life measures—United States, 1993. *Morbidity and Mortality Weekly Report, 44*, 195-200.
- Denniston, M., Brogan, D., Baker, F., Dye, J.T., & Hann, D.M. (1999, August). *Survey design challenges in sampling recently diagnosed cancer survivors*. Paper presented at the Joint Statistical Meetings, Section on Survey Research Methods, Baltimore, MD: Joint Statistical Meetings.
- Donovan, K., Sanson-Fisher, R.W., & Redman, S. (1989). Measuring quality of life in cancer patients. *Journal of Clinical Oncology, 7*, 959-968.
- Epstein, A.M., Hall, J.A., Tognetti, J., Son, L.H., & Conant, L. (1989). Using proxies to evaluate quality of life. *Medical Care, 27*, 591-598.
- Ferrell, B.R., Grant, M., Schmidt, G.M., Whitehead, C., Fonbuena, P., & Forman, S.J. (1992). The meaning of quality of life for bone marrow transplant survivors. Part 1: The impact of bone marrow transplant on quality of life. *Cancer Nursing, 15*, 153-160.
- Ferrell, B.R., Grant, M., Padilla, G., & Vemuri, S. (1991). The experience of pain and perceptions of quality of life: Validation of a conceptual model. *Hospice Journal, 7*, 9-24.
- Frank-Stromberg, M. (1984). Selecting an instrument to measure quality of life. *Oncology Nursing Forum, 11*, 88-91.
- Friedman, G.D. (1980). *Primer of epidemiology*. New York: McGraw-Hill.
- Ganz, P.A. (1994). Quality of life and the patient with cancer: Individual and policy complications. *Cancer, 74* (Suppl. 4), 1445-1452.
- Grant, M., Ferrell, B., Schmidt, G.M., Fonbuena, P., Niland, J.C., & Forman, S.J. (1992). Measurement of quality of life in bone marrow transplantation survivors. *Quality of Life Research, 1*, 375-384.
- Haberman, M.R., & Bush, N. (1998). Quality of life: Methodological measurement issues. In C.R. King, & P.S. Hinds (Eds.), *Quality of Life from nursing and patient perspectives: Theory, research, practice*. Sudbury, MA: Jones and Bartlett.
- Heiney, S.P. (1995). The healing power of story. *Oncology Nursing Forum, 22*, 899-911.
- Hinds, P.S., & Haase, J.E. (1998). Quality of life in children and adolescents with cancer. In C.R. King, & P.S. Hinds (Eds.), *Quality of life from nursing and patient perspectives: Theory, research, practice*. (pp. 93-114). Sudbury, MA: Jones and Bartlett.
- Hunt, S.M., McKenna, S.P., McEwen, J., Williams, J., & Papp, E. (1981). The Nottingham Health Profile: Subjective health status and medical consultations. *Social Science & Medicine, 15A*, 221-229.
- Johnson, J.R., & Temple, R. (1985). Food and Drug Administration requirements for approval of new anticancer drugs. *Cancer Treatment Reports, 69*, 1155-1157.
- Karnofsky, D.A., & Burchenal, J. (1949). The clinical evaluation of chemotherapeutic agents in cancer. In C.M. MacLeod (Ed.), *Evaluation of chemotherapeutic agents*. New York: Columbia Press.
- Kornblith, A.B., & Holland, J.C. (1994). *Handbook of measures for psychological, social and physical functioning in cancer* (Vol. I). New York: Memorial Sloan-Kettering Cancer Center.
- Leplege, A., & Hunt S. (1997). The problem of quality of life in medicine, *JAMA, 278*, 47-50.
- McQuellon, R.P., Russell, G.B., Cella, D.F., Craven, B.L., Brady, M., Bonomi, A., & Hurd, D.D. (1997). Quality of life measurement in bone marrow transplantation: Development of the Functional Assessment of Cancer Therapy—Bone Marrow Transplant (FACT-BMT) scale. *Bone Marrow Transplantation, 19*, 357-368.
- Meadows, A.T., Varricchio, C., Crosson, K., Harlan, L., McCormick, P., Nealon, E., Smith, M., & Ungerleider, R. (1998). Research issues in cancer survivorship. Report of a workshop sponsored by the Office of Cancer Survivorship, National Cancer Institute. *Cancer Epidemiology, Biomarkers & Prevention, 7*, 1145-1151.
- Moinpour, C.M., Feigl, P., Metch, B., Hayden, K.A., Meyskens, F.L., Jr., & Crowley, J. (1989). Quality of life end points in cancer clinical trials. *Journal of the National Cancer Institute, 81*, 485-495.
- Moinpour, C.M., Hayden, K.A., Thompson, I.M., Feigl, P., & Metch, B. (1990). Quality of life assessment in Southwest Oncology Group trials. *Oncology, 4*, 79-84, 89.

Assessing the Quality of Life of Cancer Survivors

(continued)

Mullan, F. (1985). Seasons of survival: Reflections of a physician with cancer. *New England Journal of Medicine*, 313, 270-273.

National Institutes of Health/National Heart Lung & Blood Institute. (1995) *Health-related quality of life: Findings from NHLBI-supported clinical research*. Bethesda, MD: US Department of Health and Human Services.

Osoba, D. (1994). Lessons learned from measuring health-related quality of life in oncology. *Journal of Clinical Oncology*, 12, 608-616.

Schag, C.A., Ganz, P.A., & Heinrich, R.L. (1991). Cancer Rehabilitation Evaluation System—Short Form (CARES-SF). A cancer specific rehabilitation and quality of life instrument. *Cancer*, 68, 1406-1413.

Schag, C.A., Heinrich, R.L., Aadland, R.L., & Ganz, P.A. (1990). Assessing problems of cancer patients: Psychometric properties of the Cancer Inventory of Problem Situations. *Health Psychology*, 9, 83-102.

Schipper, H. (1990). Quality of life: Principles of the clinical paradigm. *Journal of Psychosocial Oncology*, 8(2/3), 171-185.

Schipper, H., Clinch, J., McMurray, A., & Levitt, M. (1984). Measuring the quality of life of cancer patients: The Functional Living Index—Cancer: Development and validation. *Journal of Clinical Oncology*, 2, 472-483.

Spitzer, W.O., Dobson, A.J., Hall, J., Chesterman, E., Levi, J., Shepherd, R., Battista, R.N., & Catchlove, B.R. (1981). Measuring the quality of life of cancer patients: A concise QL-index for use by physicians. *Journal of Chronic Diseases*, 34, 585-597.

Sprangers, M.A.G., & Aaronson, N.K. (1992). The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: A review. *Journal of Clinical Epidemiology*, 45, 743-760.

Stewart, A.L., Hays, R.D., & Ware, J.E. (1988). The MOS Short-Form Health Survey: Reliability and validity in the patient population. *Medical Care*, 26, 724-735.

Stewart, A.L., Teno, J., Patrick, D.L., & Lynn, J. (1999). The concept of quality of life of dying persons in the context of health care. *Journal of Pain and Symptom Management*, 17, 93-108.

US Department of Health and Human Services/Public Health Service/National Institutes of Health. (1990). *Quality of life assessment in cancer clinical trials. Report of the Workshop on Quality of Life Research in Cancer Clinical Trials*. Bethesda, MD: US Department of Health and Human Services.

Visser, M.R., & Smets, E.M. (1998). Fatigue, depression and quality of life in cancer patients: How are they related? *Supportive Care in Cancer*, 101-108.

Warnecke, R.B., Ferrans, C.E., Johnson, T.P., Chapa-Resendez, G., O'Rourke, D.P., Chavez, M., Dudas, S., Smith, E.D., Martinez, S.L., Hand, R.P. & Lad, T. (1996). Measuring quality of life in culturally diverse populations. *Journal of the National Cancer Institute Monographs*, 20, 29-38.

World Health Organization—Division of Mental Health. (1993). *WHO-QOL study protocol: The development of the World Health Organization Quality of Life Assessment Instrument (MNG/PSF/93.9)*. Geneva, Switzerland: Author.

Frank Baker is the first Director of the American Cancer Society's Behavioral Research Center. The Center conducts research on behavioral and psychosocial aspects of cancer, integrates behavioral science knowledge into other ACS programs, and assists the utilization of behavioral science research in improving cancer prevention, detection, and treatment. The author of 10 books and over 100 journal articles and book chapters, Dr. Baker was on the faculties of Johns Hopkins University School of Hygiene and Public Health, the School of Medicine of the State University of New York at Buffalo, and Harvard Medical School before he came to the ACS.

The Germ

A mighty creature is the germ,
Though smaller than the pachyderm.
His customary dwelling place
Is deep within the human race.
His childish pride he often pleases
By giving people strange diseases.
Do you, my poppet, feel infirm?
You probably contain a germ.

Ogden Nash

Measurement and Methodological Issues in Minority Aging Research

Mildred Ramirez, Marvella Ford, and Anita L. Stewart

The racial, ethnic, and age composition of the U.S. population is changing rapidly. By the year 2050, nearly half will be nonwhite and 20 percent will consist of people aged 65 and older. Despite these demographic trends, ethnic minorities, older adults, women, and those with lower socioeconomic status (SES) have been underrepresented in epidemiologic and clinical studies (Ferketich, Phillips, & Verran, 1993; Forsythe & Gage, 1994; Furnham & Malik, 1994; Anderson, Aaronson, & Wilkin, 1993; Iwata & Salto, 1993). Accordingly, the National Institutes of Health now mandates the inclusion of women and minorities in research. This mandate is significant because without adequate representation of all population groups, valid generalization of results to members of these populations is not possible.

Each racial and ethnic group has unique cultural characteristics, including its own values, norms, and attitudes, (Marin et al., 1995; Marin & Perez-Stable, 1995; Ferketich et al., 1993; Devore & Schlesinger, 1987), and thus measures developed with nonminority populations may not be valid for them. Hence, it is imperative to consider, for each of these groups, whether existing measures are relevant, appropriate, reliable, and valid. Still, although the importance of cultural validity in measurement has been recognized by many researchers, (Ferketich et al.; Forsythe & Gage, 1994; Furnham & Malik, 1994; Anderson et al., 1993; Iwata & Salto, 1993), it remains common practice to apply standard measures to racial and ethnic minority groups and to lower SES groups without prior investigation of their psychometric properties for these populations.

This article describes briefly some of the issues regarding measurement in diverse groups. A more in-depth treatment of these issues as they relate to assessment of health can be found in Stewart and Napoles-Springer (in press); such measurement issues in general will be discussed at length in a forthcoming issue of the *Journal of*

Mental Health and Aging (in press) edited by Skinner, Teresi, Holmes, Stahl, and Stewart that will focus entirely on assessment in minority populations.

Resource Centers for Minority Aging Research (RCMAR)

The National Institute on Aging, the National Institute of Nursing Research, and the Office of Research in Minority Health jointly created a program (a) to decrease disparities in geriatric and gerontological research among racial, ethnic and SES groups, and (b) to increase the number of minority researchers working in these areas. Six Resource Centers for Minority Aging Research (RCMAR) and a Coordinating Center were subsequently funded to address these overarching goals (see Stahl, in press). These centers are: (a) the Center for the Active Life of Minority Elders (CALME) at Columbia Presbyterian Medical Center, New York City; (b) the Center for Aging in Diverse Communities at The University of California - San Francisco; (c) the Center on Minority Aging at the University of North Carolina - Chapel Hill; (d) the Michigan Center for Urban African American Aging Research (MCUAAR) operated jointly by Wayne State University and the University of Michigan; (e) the Native Elder Research Center at the University of Colorado Health Sciences Campus; (f) the Resource Center for African American Aging Research operated by the Henry Ford Health System in Detroit; and (g) the RCMAR Coordinating Center at the Medical University of South Carolina. While each of the Centers has its own mission and special goals, all Centers share a common goal, "...to decrease the minority/nonminority differential in health and its social sequelae for older people by focusing research upon health promotion, disease prevention, and disability prevention." (Resource Centers for Minority Aging Research, 1997).

In order to address issues specifically pertinent to measurement with older minority populations, each RCMAR has a Measurement and Methods Core, with specific goals and areas

of interest. Some of the Cores focus on both quantitative and qualitative research, while others concentrate primarily on quantitative research. Topic areas vary across sites and range from mental health to physical health, focusing, for example, on constructs such as depression, anxiety, cognition, and religiosity, and on diseases such as cancer, heart disease, and diabetes. (More specific information on the RCMARs may be found at <http://rcmar.musc.edu>.)

Sources of Cultural Bias in Measurement

Although substantial differences along various dimensions have been observed among racial, ethnic, and SES groups, (Berkanovic & Telesky, 1985; Angel & Thoits, 1987; Raczynski et al., 1994; Johnson et al., 1996; Osmond, Vranizan, Schilling, Stewart, & Bindman, 1996; Shetterley, Baxter, Mason, & Hamman, 1996), it is uncertain whether these observed differences are true differences or a result of cultural bias in the measures or methods used, (Fullerton, Wallace, & Concha-Garcia, 1993). With the increasingly widespread recognition of racial, ethnic, and class bias and insensitivity in measurement instruments and methods, there is growing demand for the validation of existing measures using samples of minority group members, and for establishing the cross-ethnic equivalence of assessment tools, (Anderson et al., 1993; Chwalow, 1995; Knight, Viridin, Ocampo, & Roosa, 1994; Bullinger, Anderson, Cella, & Aaronson, 1993; Sullivan et al., 1995). The Advisory Panel on Alzheimer's Disease, for example, specifically calls for the development and validation of screening instruments and methods that will be effective in identifying Alzheimer's across various ethno-cultural groups. (Advisory Panel on Alzheimer's Disease, 1993).

Any discussion of sources of measurement bias should include item structure, the criteria used in developing a measure, and error introduced by the interviewer and/or the respondent. (Teresi & Holmes, 1997). With regard to item structure, specific definitions and operationalizations of constructs, and the wording of particular items

may have different cultural valences across different racial, ethnic, or SES groups, that is, they may not hold the same meaning for instrument designers, raters/ interviewers, and respondents from various ethnic/racial/SES backgrounds. (Rogler, 1989). For example, many African Americans refer to diabetes as "sugar" and to hypertension as "high blood." Stevens, Kumanyika and Keil (1994) found that African American women, in response to the question of whether they were overweight, were less likely than Caucasian women to perceive themselves as being overweight despite the fact that the prevalence of obesity is twice as high among African American women as it is among Caucasian women. (Stevens et al., 1994). Cultural differences in the meaning of the term "overweight" and attitudes about the acceptability of being overweight, therefore, may well account for systematic response differences between African Americans and Caucasians to items concerning weight and body image.

Measurement bias may also be introduced in instrument administration. For example, raters/ interviewers who come from different racial/ ethnic/SES backgrounds than do the individuals being rated/interviewed may respond to cues incorrectly or in ways different than intended, or they may simply misinterpret information, leading to spurious results. (It should be noted here, too, that in addition to measurement bias, nonrepresentativeness of research samples is often cited as a factor that perpetuates racial/ ethnic/class bias in research.) (Abebimpe, 1994; Dohrenwend, 1975).

Regardless of source, measurement error, including that due to bias, has implications not only for research findings, where it might lead to erroneous results, and for epidemiological research, where measurement error may produce biased estimates of prevalence and magnitude of risk factors, but also for the development of public policies and for service delivery. Failure to account for inter- and/or intra-group differences in designing, administering, or delivering social services leads to ineffectiveness. The presumption of social or cultural homogeneity perpetuates inaccurate cultural stereotypes and thereby hinders the delivery of quality social and healthcare services

Measurement and Methodological Issues in Minority Aging Research (continued)

to racial/ethnic minorities. Lack of fit between public policy and real-world conditions, and between client needs and services delivered, are the inevitable end results where there is cultural bias in instrument design or administration, and where instruments or methods used in data gathering are insensitive to racial/ethnic differences.

Measurement Issues in Research on Aging and Minorities

Within the research community, racial/ethnic bias has been identified as a methodological issue requiring careful examination. (See Teresi & Holmes (1994) for an overview of methodological issues related to comparison of measures across subgroups.) For example, Gibson (1991), using latent variable confirmatory factor analysis, examined racial differences in the structure of and measurements made with six self-reports of health widely used in studies involving older adults. The three elements of self-reported health Gibson examined in Americans Changing Lives (House, 1986) were disease, disability, and subjective interpretation of health status. Findings showed that the form of the model had an overall acceptable fit for both the African American and Caucasian samples, indicating that, in this instance, disease, disability, and subjective interpretations of health status derive from a single latent construct, internal health state, in both groups. However, racial differences were seen in parts of the three-element model, suggesting that there are cultural and racial differences in self reports of disease, disability, and health status (Gibson, 1991). For example, subjective interpretation of health was found to be a better measure of health status for Caucasians than for African Americans, while number of chronic conditions, as an indicator of disease, was found to be a better measure of health status for African Americans than for Caucasians.

Furthermore, application of a single model to all minority populations essentially ignores not only intergroup differences, but potential intra-group variations as well (e.g., between Mexican-Americans and Puerto Ricans within the larger

Latino population, among various tribal groupings within the Native American population). Such intragroup variations have received even less attention than intergroup differences in the context of measuring health and psychosocial variables.

Some of the methodological issues related to social research in general are also applicable to minority aging research. General concerns are: errors related to items and criteria (e.g., lack of standard items with explicit coding instructions, lack of algorithms to account for missing data) (Teresi, Lawton, Ory, & Holmes, 1994), errors due to occasion (e.g., the health or psychological status of respondents or external environmental conditions); errors due to raters (e.g., reporting bias) (Teresi & Holmes, 1997), lack of established interrater reliability, lack of adequate training for assessors); and errors related or due to respondents themselves, including low level of arousal, impaired communication ability, inability to provide informed consent, decrements in vision, hearing or motor abilities, presence of conditions such as depression and fatigue, and demographic characteristics (e.g., age, gender, racial/ethnic/cultural group membership, level of education). (See Teresi & Holmes for a detailed discussion of these latter issues.)

Methods for Identifying Bias

There are three broad approaches for assessing the magnitude and nature of bias in measures across groups: qualitative studies, classic psychometric studies, and studies using modern psychometric methods.

Qualitative Studies. Qualitative studies can be used to assess the conceptual equivalence of existing measures (e.g., to explore how individuals from diverse backgrounds conceptualize a domain), and to determine whether any constructs are missing from a measurement model (Johnson et al., 1991). Qualitative approaches also can facilitate understanding of how people construct their answers -- the cognitive processes of reporting (Sudman, Bradburn, & Schwarz, 1996). Cognitive testing, focus groups, and expert panels are commonly used in qualitative measurement studies.

Classic Psychometric Studies. Applications of traditional psychometric approaches have been used to examine measures across demographic subgroups, such as comparing measurement properties across subgroups simultaneously. Classic psychometric studies include examination of methods to deal with missing data, inter- vs. intragroup variability, response bias, reliability, factor structure, content validity (a form of conceptual equivalence), and construct validity.

Modern Psychometric Methods. Because classical test theory-determined parameters and summary statistics normally vary across demographic subgroups in prevalence or distribution of items (Teresi & Holmes, 1997), more advanced psychometric methods, such as item response analysis and various types of factor analysis, are also being used to examine measurement bias. Item response theory recently has been employed to detect differential item function and item bias in epidemiological screening measures (Teresi, Cross, & Golden, 1989; Teresi & Golden, 1994; Teresi, Kleinman, & Welikson, in press), and to develop more accurate estimates of prevalence (Teresi, Albert, Holmes, & Mayeux, in press). For example, recent studies of cognitive assessment screens suggest that cognitive screens and items perform differently across groups that differ in terms of education, ethnicity and race (Albert & Teresi 1999; Teresi, 1995). In view of this, it is not surprising that research findings reflect racial, ethnic, and education subgroup differences in classification rates developed using common cognitive screening measures when such rates are compared to those provided by clinical diagnosis. (Anthony, Niaz, Le Resche, VonKorff, & Folstein, 1982; Fillenbaum, Heyman, Williams, Prosnitz, & Burchett, 1990; Escobar et al., 1986; Gurland, Wilder, Cross, Teresi, & Barnett, 1992; Valle et al., 1991).

The revelation of thought takes men out
of servitude into freedom

Ralph Waldo Emerson

Conclusions

As research increasingly takes into account, or even focuses on differences across diverse groups, issues of measurement comparability among groups are paramount. To the extent that investigators become more acquainted with these issues, there should be greater examination of measurement instruments and techniques used to improve understanding of the adequacy of existing measures and to determine the need for additional studies of measurement bias across and within cultural, racial/ethnic, and other demographically distinguishable groups.

References

- Abebimpe, V.R. (1994). Race, racism, and epidemiological surveys. *Hospital and Community Psychiatry, 45*(1), 27-31.
- Advisory Panel on Alzheimer's Disease. (1993). *Fourth Report of the Advisory Panel on Alzheimer's Disease, 1992*. NIH Pub. No. 93-3520. Washington, DC: U.S. Government Printing Office.
- Albert, S., & Teresi, J. (1999). Reading ability, education and cognitive status assessment among older adults in Harlem, New York City. *American Journal of Public Health, 89*, 95-97.
- Anderson, R.T., Aaronson, N.K., & Wilkin, D. (1993). Critical review of the international assessments of health-related quality of life. *Quality of Life Research, 2*, 365-395.
- Angel, R., & Thoits, P. (1987). The impact of culture on the cognitive structure of illness. *Culture, Medicine, and Psychiatry, 11*, 465.
- Anthony, J.C., Niaz, U., LeResche, L.A., VonKorff, M.R., & Folstein, M.F. (1982). Limits of the Mini-Mental State as a screening test for dementia and delirium among hospital patients. *Psychological Medicine, 12*, 397-408.
- Arean, P.A., & Gallagher-Thompson, D. (1996). Issues and recommendations for the recruitment and retention of older ethnic minority adults into clinical research. *Journal of Consulting and Clinical Psychology, 64*, 878.
- Berkanovic, E., & Telesky, C. (1985). Mexican-American, Black-American and White-American differences in reporting illnesses, disability, and physician visits for illnesses. *Social Science & Medicine, 20*, 567.
- Bullinger, M., Anderson, R., Cella, D., & Aaronson, N. (1993). Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. *Quality of Life Research, 2*, 451-459.

Measurement and Methodological Issues in Minority Aging Research (continued)

Chwalow, A.J. (1995). Cross-cultural validation of existing quality of life scales. *Patient Education & Counseling*, 26(1-3), 313-318.

Devore, W., & Schlesinger, E.G. (1987). *Ethnic-sensitive social work practice* (2nd ed.). Columbus, OH: Merrill Publishing.

Dohrenwend, B.P. (1975). Sociocultural and social-psychological factors in the genesis of mental disorders. *Journal of Health and Social Behavior*, 16, 365-392.

Escobar, J., Burnam, M., Karno, M., Forsythe, A., Landsverk, J., & Golding, J.M. (1986). Use of the Mini-Mental State Examination (MMSE) in a community population of mixed ethnicity. *Journal of Nervous and Mental Disease*, 174, 607-614.

Ferketich, S., Phillips, L., & Verran, J. (1993). Focus on psychometrics: Development and administration of a survey instrument for cross-cultural research. *Research in Nursing & Health*, 16, 227-230.

Fillenbaum, G.G., Heyman, A., Williams, K., Prosnitz, B., & Burchett, B. (1990). Sensitivity and specificity of standardized screens of cognitive impairment and dementia among elderly Black and White community residents. *Journal of Clinical Epidemiology*, 43, 650-660.

Forsythe, H.E., & Gage, B. (1994). Use of a multicultural food-frequency questionnaire with pregnant and lactating women. *American Journal of Clinical Nutrition*, 59, 203S-206S.

Fullerton, J.T., Wallace, H.M., & Concha-Garcia, S. (1993). Development and translation of an English-Spanish dual-language instrument addressing access to prenatal care for the border-dwelling Hispanic women of San Diego County. *Journal of Nurse-Midwifery*, 38, 45-49.

Furnham, A., & Malik, R. (1994). Cross-cultural beliefs about depression. *International Journal of Social Psychiatry*, 40, 106-123.

Garland, B.J., Wilder, D.E., Cross, P.E., Teresi, J.A., & Barnett, V.W. (1992). Screening scales for dementia: toward reconciliation of conflicting cross-cultural findings. *International Journal of Geriatric Psychiatry*, 7, 105-113.

Gibson, R.C. (1991). Race and the self-reported health of elderly persons. *Journal of Gerontology: Social Sciences*, 46, S235-S242.

House, J.S. (1986, Wave 1). Americans' changing lives. [Computer file]. Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

Iwata, N., & Salto, K. (1993). The factor structure of the 28-item General Health Questionnaire when used in Japanese early adolescents and adult employees: Age- and cross-cultural comparisons. *European Archives of Psychiatry and Clinical Neuroscience*, 242, 172-178.

Johnson, T.P., O'Rourke, D., Chavez, N., et al. (1996). Cultural variations in the interpretation of health survey questions. In R. Warnecke (Ed.), *Health survey research methods*. Hyattsville, MD: National Center for Health Statistics.

Kaye, J.M., Lawton, P., & Kaye, D. (1990). Attitudes of elderly people about clinical research on aging. *Gerontologist*, 30, 100.

Knight, G.P., Virdin, L.M., Ocampo, K.A., & Roosa, M. (1994). An examination of the cross-ethnic equivalence of measures of negative life events and mental health among Hispanics and Anglo-American children. *American Journal of Community Psychology*, 22(5), 767-783.

Marin, G., Burhansstipanov, L., Connell, C.M., Gielen, A.C., Helitzer-Allen, D., Lorig, K., Morisky, D.E., Tenney, M., & Thomas, S. (1995). A research agenda for health education among underserved populations. *Health Education Quarterly*, 22, 346-363.

Marin, G., & Perez-Stable, E.J. (1995). Effectiveness of disseminating culturally appropriate smoking-cessation information: Programa Latino Para Dejar de Fumar. *Journal of the National Cancer Institute Monographs*, 18, 155-163.

McCarthy, C.R. (1994). Historical background of clinical trials involving women and minorities. *Academic Medicine*, 69, 695.

Osmond, D.H., Vranizan, K., Schillinger, D., Stewart, A.L., & Bindman, A.B. (1996). Measuring the need for medical care in an ethnically diverse population. *Health Services Research*, 31, 551.

Raczynski, J.M., Taylor, H., Cutter, G., Hardin, M., Rappaport, N., & Oberman, A. (1994). Diagnoses, symptoms, and attribution of symptoms among Black and White inpatients admitted for coronary heart disease. *American Journal of Public Health*, 84, 951.

Resource Centers for Minority Aging Research. (1997). *NIH Guide*, 26, (1), January 10.

Rogler, L.H. (1989). The meaning of culturally sensitive research in mental health. *American Journal of Psychiatry*, 146, 296.

Shetterly, S.M., Baxter, J., Mason, L.D., & Hamman, R.F. (1996). Self-rated health among Hispanic vs non-Hispanic White adults. The San Luis Valley Health and Aging Study. *American Journal of Public Health*, 86, 1798.

Measurement and Methodological Issues in Minority Aging Research (continued)

Stahl, S.M. (in press). A long-range innovative approach to reducing health disparities: The National Institute on Aging's Resource Centers for Minority Aging Research. In J.H. Skinner, J.A. Teresi, D. Holmes, S.M. Stahl, & A.L. Stewart (Eds.). (in press). Measurement in older ethnically diverse populations. *Journal of Mental Health and Aging*.

Stevens, J., Kumanyika, S.K., & Keil, J.E. (1994). Attitudes toward body size and dieting: Differences between elderly Black and White women. *American Journal of Public Health*, 84, 1322-1325.

Stewart, A.L., & Napoles-Springer, A. (in press). Health related quality of life assessments in diverse population groups in the United States. *Medical Care*.

Sudman, S., Bradburn, N.M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.

Sullivan, L.M., Dukes, K.A., Harris, L., Dimus, R.S., Greenfield, S., & Kaplan, S.H. (1995). A comparison of various methods of collecting self-reported health outcomes data among low-income and minority patients. *Medical Care*, 33, AS183-AS194.

Swanson, G.M., & Ward, A.J. (1995). Recruiting minorities into clinical trials: Toward a participant-friendly system. *Journal of the National Cancer Institute*, 87, 1747.

Teresi, J.A., Cross, P.S., & Golden, R.R. (1989). Some applications of latent trait analysis to the measurement of ADL. *Journal of Gerontology*, 44, S196-S204.

Teresi, J.A., & Golden, R.R. (1994). Latent structure methods for estimating item bias, item validity and prevalence using cognitive and other geriatric screening measures. *Alzheimer Disease and Associated Disorders*, 8, S291-S298.

Teresi, J.A., & Holmes, D. (1994). Overview of methodological issues in gerontological and geriatric measurement. In M.P. Lawton & J.A. Teresi (Eds.), *Annual review of gerontology and geriatrics: Focus on assessment techniques*, Vol 14, New York: Springer.

Teresi, J.A., Lawton, M.P., Ory, M., & Holmes, D. (1994). Measurement issues on chronic care populations: Dementia special care. *Alzheimer Disease and Associated Disorders*, 8(Supplement 1), S144-S183.

Teresi, J.A., Golden, R.R., Cross, P., Gurland, B., Kleinman, M., & Wilder, D. (1995). Item bias in cognitive screening measures: Comparisons of elderly White, Afro-American, Hispanic and high and low education subgroups. *Journal of Clinical Epidemiology*, 48, 473-483.

Teresi, J.A., & Holmes, D. (1997). Methodological issues in cognitive assessment and their impact on outcome measurement. *Alzheimer Disease and Associated Disorders*, II(Supplement 6), 146-155.

Teresi, J.A. & Holmes, D. (1997). Reporting source bias in estimating prevalence of cognitive impairment. *Journal of Clinical Epidemiology*, 50, 175-184.

Teresi, J., Albert, S., Holmes, D., & Mayeux, R. (in press). Use of latent class analyses for the estimation of prevalence of cognitive impairment, and signs of stroke and Parkinson's disease among African-American elderly in Central Harlem: Results of the Harlem Aging Project. *Neuroepidemiology*.

Teresi, J., Kleinman, M., & Welikson, K. (in press). Modern psychometric methods for detection of differential item functioning: Application to cognitive assessment measures. *Statistics in Medicine*.

Valle, R., Hough, R., Kolody, B., Cook-Gait, H., Velazquez, G.F., Jimerez, R., Atkinson, J.H., Jex, N., Need, D., Mendivil, L., Sanchez, C., Lopez, S., Lujan, G., Ballon, E., & Ganger, W. (1991). The validation of the Blessed Mental Status Test and the Mini Mental Status Examination with a Hispanic population. In *Final report to the National Institute of Mental Health, The Hispanic Alzheimer's Research Project (HARP)*. San Diego, CA: San Diego State University.

Vernon, S.W., Roberts, R.E., & Lee, E.S. (1984). Ethnic status and participation in longitudinal health surveys. *American Journal of Epidemiology*, 119, 99.

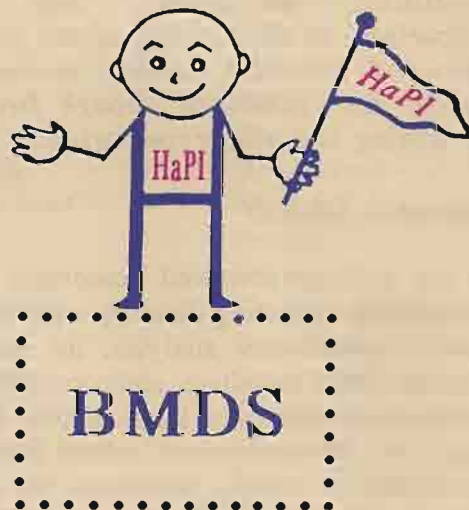
Mildred Ramirez is an Associate in the Research Division of the Hebrew Home for the Aged in Riverdale, NY, and Director of the Measurement and Methods Core of the NIA-funded Columbia [University] Center for the Active Life of Minority Elders. Dr. Ramirez's research interests include cultural diversity in nursing homes in the context of quality of care and job-related outcomes, relationships between race and institutional risk, and interrater reliability of measures across different groups.

Marvella Ford is an Associate Research Scientist with the Henry Ford Health System Center (HFHS) for Research in Diverse Populations (CRDP), and Director of the Measurement and Methods Core of the NIA-funded Resource Center for African American Aging Research of the CRDP.

Dr. Ford's work is focused in the areas of psychosocial factors affecting the use of health services by African Americans, and the development of culturally relevant data collection instruments and outcome measures.

Anita L. Stewart is Professor in the Institute for Health and Aging at the University of California at San Francisco, Co-Director of UCSF's NIA-funded Resource Center on Minority Aging Research, and Director of the Center's Measurement and Methods Core. Her research interests encompass the conceptualization and measurement of health-relevant concepts, measurement of health-relevant factors in diverse populations, and adapting existing measures for use with minority populations. Dr. Stewart was a major developer of the long-term health measures of the Medical Outcomes Study.

Remember me,
I'm HaPI at
BMDS!



Search for measurement instruments
in the Health and Psychosocial Instruments (HaPI)
database with 95,000 records of measurement
instruments online or on CD-ROM!



Produced by BMDS - Behavioral Measurement Database Services

SUBSCRIBE TO HaPI!



Call Ovid Technologies at 800-950-2035 for pricing
and to order today!

Obtain copies of measurement instruments
from BMDS Instrument Delivery!



Call BMDS at 412-687-6850, fax 412-687-5213, or
e-mail bmdshapi@aol.com (\$20/\$30 processing fee)

Culture . . . One Step at a Time (Part 2)

John B. Gatewood

Editor's Introduction to Part 2: In the first part of Dr. Gatewood's piece, published in the previous issue of The BML (6(2):5-10), he discussed three techniques by which one can begin to gain an understanding of a culture -- free-listing, pile-sorting, and triadic comparison. In this second of two parts, Dr. Gatewood discusses various techniques to analyze data generated through free-listing, pile-sorting, and triadic comparison.

Consensus Analysis

So far, I have reviewed somewhat unusual methods for collecting data. By contrast, I now consider consensus analysis, an interesting technique that cognitive anthropologists have devised to analyze these kinds of data. Although many of the ingredient ideas have been percolating in various literatures for decades, Romney, Weller, and Batchelder (1986) were the first to propose the formal theory and mathematical procedures of consensus analysis. Here I can do no more than sketch the basics, and I must refer readers interested in more details to the original source.

The essential problem that consensus theory addresses is as follows. Given that members of a culture do not uniformly agree with one another in their beliefs about what is true or proper, how can an outsider tell if there is a 'common culture' underlying their diverse opinions? The key to answering this question lies in realizing that (a) no one knows all of his or her group's culture and (b) agreement is a matter of degree. In particular, experts in a cultural domain should agree with one another more than nonexperts do (see Boster, 1985). Following this intuition, consensus theory assumes that "the correspondence between the answers of any two informants is a function of the extent to which each is correlated with the truth" (Romney et al., 1986, p. 316). Consensus theory focuses precisely on the extent to which informants converge on the same answers to systematically asked questions as a measure of cultural knowledge.

For example, suppose Mr. Smith gives a multiple-choice test to his class, but arriving home discovered that he has lost the answer key. Could he grade the students' answer sheets anyway? Yes, he could (Batchelder & Romney, 1988). According to consensus theory, if students did not know the correct answer to a question, then they would just guess, and such guessing should produce predictable proportions of agreement across the available answers. On the other hand, if students know the correct answer, then they will converge on the same answer (the "correct" one) more frequently than expected just by chance. Knowledge -- cultural competence in a domain -- produces deviations from equiprobability, and more knowledgeable individuals will agree with one another more often than less knowledgeable individuals do.

The ingenuity of consensus analysis is that it provides a way to estimate the cultural competence of individual informants from the patterning of their agreement. The formal model rests on three central assumptions (Romney et al., 1986, pp. 317-318):

1. *Common Truth.* The informants all come from a common culture, such that whatever their cultural version of the truth is, it is the same for all informants.
2. *Local Independence.* Informants' answers are given independently of other informants, i.e., there is no collusion or influence among informants.
3. *Homogeneity of Items.* Questions are all of the same difficulty, such that the informant's cultural competence is equal across all questions.

Certain kinds of statistical criteria serve as checks on the validity of these three critical assumptions, though a detailed description of these technicalities is beyond the scope of this column. If these technical criteria are met, then it is considered reasonable to compute a relative "competence score" for each informant, as a measure of how well that particular individual represents the entire sample's answers to the questions asked. Under these conditions, a competence score can be interpreted as the proportion of questions an informant answered "correctly." Conversely, if these statistical checks on the validity of underlying assumptions are violated, then one or more of

the three critical assumptions must *not* be true of the data. For example, it may well be that there is no common culture within the sample, but rather many different subcultures, or systematically different ways of responding.

Consensus analysis works well with many kinds of data: true-false, check lists, belief-frames, multiple-choice, rankings, ratings, and even proximity matrices (such as similarity matrices). There are, however, two important limitations: (a) the battery of questions must be of a single type, such as all multiple-choice, all similarity matrices, etc., and (b) the questions must ask informants for conventional truths or judgments, *not* their personal preferences or histories. For instance, consensus analysis is well-suited for questions asking informants to check all diseases on a list that are serious, contagious, result in a high fever, or for which one should see a doctor. But consensus analysis makes little sense if informants are asked to check all the diseases that they have actually had in their lives. Likewise, consensus analysis is appropriate for pile-sort data of mixed drinks based on their similarities, but not for pile-sorts based on whether the drinks taste good. The reason is simple: agreement with respect to preferences or histories is not knowledge-driven.

An Integrative Example

Results from a small class project will illustrate how these different methods and analyses can be used in concert. The data come from 14 college students, whose knowledge of a single cultural domain (kinds of fish) was probed via free-listing, single pile-sort, triadic comparison, and several ranking and rating tasks.

Free-listing was the first task. The instructions were to list all the different kinds of fish they could think of in 90 seconds. Collectively, the 14 informants produced lists containing 226 items totaling to 115 different kinds of fish. Table 6 shows the aggregated results for the 43 kinds of fish mentioned by at least two people.

Table 6
The 43 Most Frequently Listed Kinds of Fish
(N=14)

Goldfish	10	Steelhead	2
Trout	8	Rock bass	2
Salmon	8	Beta	2
Flounder	7	Great white shark	2
Catfish	7	Bullhead catfish	2
Sunfish	7	Brown trout	2
Bass	7	Brook trout	2
Tuna	7	Pink salmon	2
Piranha	6	Barracuda	2
Shark	5	Bluegill	2
Swordfish	5	Striped bass	2
Rainbow trout	5	Blue catfish	2
Carp	4	Remora	2
Guppy	4	Jellyfish	2
Cod	3	Channel catfish	2
Muskie	3	Dog salmon	2
Red salmon	3	Smallmouth bass	2
Bluefish	3	Clams	2
Largemouth bass	3	Minnow	2
Pickering	3	Yellowfin tuna	2
Blowfish	2	Scrod	2
Walleye	2		

From the 115 kinds of fish the students free-listed, I chose 19 varieties for further study. Given the large number of items in this cultural domain, single pile-sorting was the most obvious method for obtaining similarity judgments. I could have chosen more items for this task, but I wanted students to see that similar results could be obtained via triadic comparisons. By using a balanced incomplete block design for the triads task and the same items for the pile-sort, the results of both methods would be directly comparable without overburdening my captive students. Thus, students did the single pile-sort task one day and the triadic comparisons the following class period.

Figure 1 is a nonmetric multidimensional scaling of the aggregate similarity matrix from the single pile-sort data, and Figure 2 is from the triads data. (In these plots, the closer items appear to one another, the more similar students judged them to be. The items are displayed in relation to horizontal and vertical axes, whose underlying meaning the researcher must interpret subjectively. "Stress," a measure of

how well the particular configuration fits the data, ranges from 0-1, with lower values reflecting better fit.) Visual comparison of the two figures indicates the two methods produced similar results. Indeed, the Pearson r between the two similarity matrices is .74. Normally, we would expect to find greater reliability between these two methods, but bear in mind that we had only 14 students doing the single pile-sort rather than the recommended 30-40. Also, when implementing the BIB (balanced incomplete block design) triads design, I chose to give each student the same subset of triads (so I could better compare informants), rather than using different instances of the design for each student (the domain-focused approach). Given these shortcomings of the class project, the obtained reliability of .74 is not bad.

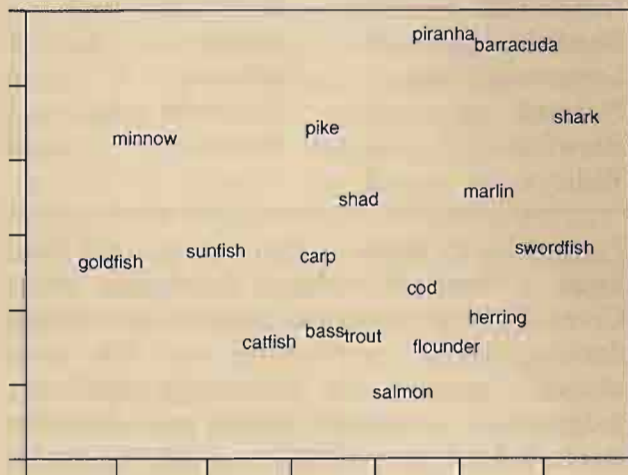


Figure 1. MDS from Single Pile-Sorts (stress = .15)

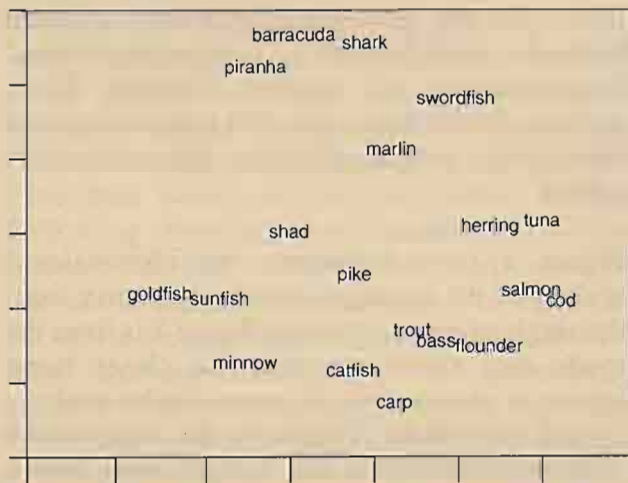


Figure 2. MDS from Triadic Comparisons (stress = .20)

Because I was most interested in using a variety of methods to compare informants with respect to their knowledge of fish, I included three auxiliary rating and ranking tasks after we had completed the free-listing, pile-sort, and triadic comparisons. The first of these asked informants to rate their own knowledge of each of the 19 fish varieties using a 5-point scale, where 1 indicated "never heard of it before" and 5 indicated "know quite a bit." And, because there had been some class discussion concerning kinds of fish following the pile-sort and triads tasks, the second and third auxiliary tasks asked each informant to rate all the students on a 5-point "novice to expert" scale, then to rank order everyone in the class with respect to how much they knew about fish.

In all, the project garnered four direct measures of informants' knowledge of the domain (length of free-list, self-rating of knowledge, social-rating of expertise, and social-ranking of expertise) and three indirect measures (competence scores from consensus analyses). Table 7 presents these informant-level measures.

Note that Table 7 displays three competence scores: one for the pile-sort data, and two for the triads data. This is because there are two ways to handle consensus analysis for the triads task. One can use either (a) the lower half of each individual's similarity matrix, as is done for pile-sort data (resulting in $n(n-1) / 2$, or 171, "multiple-choice" items), or (b) the non-randomized tallies of the triadic sets that were actually given to each informant (114 "multiple-choice" items, given the BIB design that I used).

Incidentally, all three consensus analyses meet the statistical criteria of the formal model. Hence, the patterning of agreement among the 14 students indicates there is a culturally correct way of doing the pile-sort, and likewise for the triads task. Kevin best exemplifies the sample's "correct" way of pile-sorting (competence score of .89), whereas Jennifer (.67 and .68) and Lisa (.66 and .69) best represent the sample's "correct" responses to the triads task.

The last points I will make from this example concern the correlations among different individual-level measures. Table 8 shows the correlation matrix of all seven measures, which includes a couple of surprising findings. (Again,

Table 7
Informant-Level Measures

Students	Items in Free-List	Mean Self-Rating of 19 fish	Mean Soc-Rating of Expertise	Mean* Soc-Ranking of Expertise	Pile-Sort Comp. Score	Triads Comp. Score	Triads Comp. Score
Josh	34	4.95	5.00	14.00	.78	.43	.38
Matt	30	3.84	3.50	10.00	.32	.41	.37
Sara	26	3.05	2.64	7.79	.80	.62	.63
Judy	21	3.68	3.00	8.07	.74	.61	.62
Corey	19	3.47	3.07	8.21	.55	.56	.63
Hassan	17	4.16	3.93	11.50	.82	.37	.24
Nora	15	2.95	2.43	6.00	.25	.57	.61
Jennifer	13	2.53	1.86	3.71	.03	.67	.68
Kevin	12	3.37	3.50	10.07	.89	.56	.59
Derek	10	4.00	3.43	10.14	.77	.47	.46
Lisa	10	2.84	1.86	2.36	.70	.66	.69
Beth	7	2.79	1.93	4.50	.74	.52	.53
Serene	7	2.84	1.64	3.14	.68	.62	.64
Hope	5	2.58	2.14	5.50	.43	.63	.67
M	16.14	3.36	2.85	7.50	.61	.55	.55
SD	8.60	.67	.93	3.32	.25	.09	.13

Note: *The scores for expertise ranking (fourth column) have been inverted to keep the meaning of correlation coefficients' signs clear in what follows. For example, Josh was uniformly regarded as the most knowledgeable in this domain; hence, his inverted score is 14 rather than 1.

note that we display only the lower half of the matrix because it is symmetric.)

The seven measures form into three logical groupings. First, length of free-list and self-rating of knowledge with respect to the 19 fish are both related to how much informants really know about this cultural domain. But as Don Campbell was fond of saying, "All measures are fallible"; and the correlation between these two is somewhat lower than one would like ($r = .68$). Length of one's free-list is affected by motivation, and self-ratings presume shared understanding of the response scale. Second, the two social evaluations differ only in the manner in which informants are allowed to express their opinions -- rating versus ranking -- and are highly correlated ($r = .98$). And third, because the instructions for both the pile-sorting and triadic comparisons tasks asked for overall similarity judgments, we might expect all three competence scores to be positively interrelated. But as inspection of Table 8 reveals, this is not the case.

The two ways of assessing triadic competence are highly correlated ($r = .97$), as expected, but competence doing the pile-sort task is largely independent of these two measures, and even

tends to be *inversely* related to triadic competence ($r = -.27$ and $-.28$, respectively). How can this be? What does it mean? Recall that the two methods did produce convergent results with respect to their aggregate item-by-item similarity matrices (see Figures 1 and 2); hence, the negative correlation does not bear on the question of intermethod reliability. Rather, competence scores indicate the degree to which individual's responses represent the patterning of agreement in the entire sample -- how well an individual "measures" the group consensus. Thus, the "surprising" negative correlation simply means that students whose pile-sorts resembled those of other students tended to be more idiosyncratic in the ways they answered the triadic comparisons, and vice versa.

Another interesting finding in Table 8 is the ability of students to evaluate one another's expertise. That is, the correlations between the social evaluations of expertise and the self-generated indicators of knowledge (i.e., free-listing and self-rating) are quite high ($r_s = .67 - .95$). Having been in the classroom and observed the brief occasions when anyone demonstrated knowledge or lack thereof, I marvel at the students' sensitivity to and convergent interpretations of very subtle social clues.

Lastly, perhaps the most puzzling and important finding in Table 8 is the lack of correspondence between the direct measures of informants' knowledge of fish and their cultural competence scores on the pile-sort and triads tasks. Indeed, the correlations between the two types of measures are either nonsignificant or actually negative. The more knowledgeable informants (as determined by the four direct measures) are slightly more typical of the group with respect to their pile-sorts. But contrary to other studies (Brewer, 1995), they are very *atypical* with respect to their triadic judgments. Indeed, for the triads task, the most domain-knowledgeable informants are actually the poorest representatives of the sample's common culture. Their greater knowledge of fish did *not* produce

greater agreement; rather the sample's consensus was formed by relatively ignorant informants who agreed among themselves.

This latter finding provides a general caution to those who would use consensus analysis uncritically. There are situations where 'knowledge of the common culture' means being fairly ignorant, and sometimes ignorance produces its own patterning of agreement. For some tasks, less knowledgeable people see only one way of responding, whereas experts are aware of alternatives (see also, Boster & Johnson, 1989). In such cases, the more informants know, the less likely they are to agree with others. As a way of reminding ourselves that such possibilities exist, it might be wise to relabel "competence scores" as "representative scores" in conducting consensus analysis.

Table 8
Pearson Correlation Coefficients Among Informant-Level Measures

	Free-List	Self-Rating of 19 Fish	Soc-Rating of Expertise	Soc-Ranking of Expertise	P-S: Comp	Tri: Comp. 1
Self-Rating	.68	--	--	--	--	--
Soc-Rating	.70	.95	--	--	--	--
Soc-Ranking	.68	.91	.98	--	--	--
P-S: Comp.	.04	.43	.38	.38	--	--
Tri: Comp. 1	-.46	-.80	-.78	-.79	-.27	--
Tri: Comp 2	-.46	-.79	-.75	-.76	-.28	.97

Note: **Boldfaced** values are significant at $p \leq .05$, two-tailed

References

- Batchelder, W., & Romney, A.K. (1988). Test theory without an answer key. *Psychometrika*, 53, 71-92.
- Boster, J.S. (1985). "Requiem for the omniscient informant": There's life in the old girl yet. In J. Dougherty (Ed.), *Directions in cognitive anthropology* (pp. 177-197). Urbana, IL: University of Illinois Press.
- Boster, J.S., & Johnson, J.C. (1989). Form or function: A comparison of expert and novice judgments of similarity among fish. *American Anthropologist*, 91, 866-889.
- Brewer, D.D. (1995). Cognitive indicators of knowledge in semantic domains. *Journal of Quantitative Anthropology*, 5, 107-128.
- Romney, A.K., Weller, S.C., & Batchelder, W.H. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist*, 88, 313-338.

John B. Gatewood is Professor of Anthropology at Lehigh University. He has roughly 35 professional publications in the areas of cognitive anthropology, fisheries ethnography, tourism studies, and linguistic anthropology. In addition, Dr. Gatewood has considerable experience doing marketing and advertising research for a variety of firms. His current research interests include distributive models of culture, network analysis of organizations, motivations of heritage tourists, and theoretical issues concerning the 'units' of culture.

HaPI Advisory Board

Aaron T. Beck, MD
University of Pennsylvania School of Medicine

Timothy C. Brock, PhD
Ohio State University, Psychology

William C. Byham, PhD
Development Dimensions International

Donald Egolf, PhD
University of Pittsburgh, Communication

Sandra J. Frawley, PhD
Yale University School of Medicine, Medical Informatics

David F. Gillespie, PhD
*Washington University
George Warren Brown School of Social Work*

Robert C. Like, MD, MS
*University of Medicine and Dentistry of New Jersey,
Robert Wood Johnson Medical School*

Joseph D. Matarazzo, PhD
Oregon Health Sciences University

Vickie M. Mays, PhD
University of California at Los Angeles, Psychology

Michael S. Pallak, PhD
Behavioral Health Foundation

Kay Pool, President
Pool, Heller & Milne, Inc.

Ellen B. Rudy, PhD, RN, FAAN
University of Pittsburgh School of Nursing

Gerald Zaltman, PhD
*Harvard University Graduate School of
Business Administration*

Stephen J. Zyzanski, PhD
Case Western Reserve University School of Medicine

HaPI Thoughts

That sounds great, Brad, but why do you sing just one measure of a song and stop?



*Oh, my psychology professor asked me to choose an appropriate **measure** for a research project!*



Are you mixed-up about measures? Please note: BMDS and HaPI can help sharpen your measurement repertoire. We'll help you to find an appropriate measurement instrument for virtually any need. Call BMDS today!

DOUBLE ISSUE

The **BEHAVIORAL MEASUREMENT** *Letter*

**Behavioral
Measurement
Database
Services**

In This Issue:

- Introduction to This Issue
Al K. DeRoy 1
- Assessing the Quality of Life
of Cancer Survivors
Frank Baker 2
- Measurement and Methodological
Issues in Minority Aging Research
*Mildred Ramirez, Marvella Ford, and
Anita L. Stewart* 13
- Culture . . . One Step at a Time (Part 2)
John B. Gatewood 20
- HaPI Thoughts 25

The Behavioral Measurement Letter
BIMDS
Behavioral
Measurement
Database
Services
PO Box 110287 • Pittsburgh, PA 15232-0787
DATED MATERIAL

