Measurement Issues in Health Disparities Research

*Mildred Ramírez, Marvella E. Ford, Anita L. Stewart,
and Jeanne A. Teresi*

Background. Racial and ethnic disparities in health and health care have been documented; the elimination of such disparities is currently part of a national agenda. In order to meet this national objective, it is necessary that measures identify accurately the true prevalence of the construct of interest across diverse groups. Measurement error might lead to biased results, e.g., estimates of prevalence, magnitude of risks, and differences in mean scores. Addressing measurement issues in the assessment of health status may contribute to a better understanding of health issues in cross-cultural research.

Objective. To provide a brief overview of issues regarding measurement in diverse populations.

Findings. Approaches used to assess the magnitude and nature of bias in measures when applied to diverse groups include qualitative analyses, classic psychometric studies, as well as more modern psychometric methods. These approaches should be applied sequentially, and/or iteratively during the development of measures.

Conclusions. Investigators performing comparative studies face the challenge of addressing measurement equivalence, crucial for obtaining accurate results in cross-cultural comparisons.

Key Words. Measurement, cross-cultural, health disparities

The proportion of both minority and older adults in the United States is growing; as a result, the older population is becoming more racially and ethnically diverse (Ford and Hatchett 2001; Sinclair et al. 2002; Federal Interagency Forum on Aging-Related Statistics 2004). Members of minority groups have higher rates of morbidity and mortality than do their counterparts in the general population for almost all categories of disease (Sinclair et al. 2002; Federal Interagency Forum on Aging-Related Statistics 2004; Frist 2005). Racial and ethnic disparities in health and health care have been well documented, and the elimination of such disparities is currently a part of a national agenda (Fiscella et al. 2000; Ashton et al. 2003). It is, therefore, not surprising

that such racial and ethnic disparities in health are also reflected within the U.S. veteran population (Young, Maynard, and Boyko 2003; Zingmond et al. 2003), a focus of this special issue. Thus, addressing measurement issues in the assessment of the health status of diverse populations of older adults is of critical importance. Measurement accuracy (or inaccuracy) can affect study results producing, for example, biased estimates of symptoms and disorder, and the generation of misleading conclusions. This article provides a brief overview of the issues regarding measurement in diverse populations.

The physical health challenges facing older racial and ethnic minority group members are numerous. Data show, for instance, that older African Americans, compared with older whites, have a higher incidence of hypertension, heart disease, stroke, and end-stage renal disease (Kotchen et al. 1998; Sinclair et al. 2002). In fact, the prevalence of hypertension is 50 percent higher in African American than in white adults (Kotchen et al. 1998). Additionally, African American men have both the highest incidence of and associated mortality from prostate cancer than any other racial or ethnic group, and this disparity continues to increase (Guo et al. 2000; Powell et al. 2000). Older Latinos, like older African-American adults, appear to have worse physical health than do older white adults (Villa and Aranda 2000; American Heart Association National Center 2005; National Institute of Diabetes and Digestive and Kidney Diseases 2005). Latinos are 1.9 times more likely to have diabetes than are whites of similar age; 25–30 percent of Latinos age 50 years or older have either diagnosed or undiagnosed diabetes (National Institute of Diabetes and Digestive and Kidney Diseases 2005). Additionally, Latinos, like African Americans, appear to be at particular risk for cardiovascular disease and stroke, which account for 31 percent of all Latino deaths annually (American Heart Association National Center 2005). Asian American women, in particular, appear to be at greater risk for breast cancer than are other women; breast cancer is the most common cause of cancer incidence and mortality among members of this group (Kagawa-Singer and Pourat 2000; Tanjasiri and Sablan-Santos 2001).

Address correspondence to Mildred Ramírez, Ph.D., Research Division, The Hebrew Home for the Aged at Riverdale, 5901 Palisade Avenue, Riverdale, NY 10471. Marvella Ford, Ph.D., is with the Department of Biostatistics, Bioinformatics, and Epidemiology, Medical University of South Carolina, Charleston, SC. Anita Stewart, Ph.D., Professor, is with the Institute for Health & Aging, University of California–San Francisco, San Francisco, CA. Jeanne Teresi, Ed.D., Ph.D., is with the Columbia University Stroud Center, Faculty of Medicine, and the New York State Psychiatric Institute, The Hebrew Home for the Aged at Riverdale, Research Division, Riverdale, NY.

Some studies have shown that health-related disparities among racial and ethnic groups disappear or are attenuated once confounding demographic variables such as income and education have been controlled (de Rekeneire et al. 2003; Bromberger et al. 2004). However, a greater number of studies demonstrate that racial and ethnic disparities remain even when such adjustments have been implemented (see Mayberry, Mili, and Ofili 2000; Kressin and Petersen 2001). Racial and ethnic disparities continue to be observed in epidemiological research, as reflected in different levels of risk factors, dissimilar rates of disease, differing responses to treatment, and unequal quality and access to care (Schneider, Zaslavsky, and Epstein 2002; Smedly, Stith, and Nelson 2002).

Racial and ethnic disparities in mortality rates may be because of comorbidity, access to health services, knowledge, attitudes and beliefs about disease, and/or disease biology (Ford and Hatchett 2001; Kaplan and Bennett 2003; Sankar et al. 2004). Longitudinal studies identifying other factors associated with disparities are needed because causal relationships involving health disparities and demographic factors cannot be determined from cross-sectional analyses, such as those presented in many of the studies cited above. However, a prerequisite to the alleviation of health disparities among racially diverse populations is addressing possible measurement bias in the assessment of self-reported health status. That is, in order for cross-cultural research to be conducted in a meaningful manner, it is important to determine first whether measures developed among nonminority populations perform in the same way when applied to minority populations.

Each racial and ethnic group has unique cultural characteristics including values, norms, and attitudes (Mutran, Reed, and Sudha 2001; Napoles-Springer and Stewart 2001; Shire 2002; Cabassa 2003). Hence, it is imperative to consider, for each of these groups, whether existing measures are relevant, appropriate, reliable, and valid. Although the importance of the cultural validity of questionnaire items has been recognized by many researchers (Angel and Frisco 2001; Mui, Burnette, and Chen 2001; Napoles-Springer and Stewart 2001), the practice of applying standard measures to groups of racial and ethnic minorities (Robin et al. 2003; Stanley and Chang 2003), and to groups with lower socioeconomic status without investigation of the psychometric properties for these populations remains common.

The development of culturally equivalent measures represents a step forward in the accurate assessment of health, health determinants and outcomes in the context of multicultural research, thus potentially contributing to the alleviation of health disparities. Increasing attention is being paid to the

measurement of physical and mental health constructs in different racial and ethnic groups. Following is a brief overview of some of the issues regarding measurement in diverse populations.

IMPORTANCE OF KNOWING WHETHER EXISTING MEASURES ARE ADEQUATE ACROSS CULTURAL AND AGE SUBGROUPS

Substantial differences related to physical health and mental health outcomes have been observed across different ethnic/racial groups (Neighbors et al. 2003; Turner and Avison 2003; Cohen et al. 2004). However, it is uncertain whether these observed differences reflect true differences, or whether they merely reflect cultural bias in the measures (Snowden 2003; Scholderer, Grunert, and Brunso 2005).

Measurement error can occur both through cross-cultural differences in the interpretations of the meaning of concepts and of items used to measure constructs (Gannotti et al. 2001; Moors 2004). The definition and operationalization of constructs, as well as the selection of items are likely to have different cultural meaning or value, and might reflect the idiosyncrasies of a particular societal group (Gierl 2000; Trinh, Ngo, and Conway 2003). Individuals from different cultures and physical environments are likely to have experienced differences in their cognitive and perceptual development. It may be unrealistic to assume automatically that concepts can be measured in the same way for all groups of people. That is, an item might not have the same meaning for either raters/interviewers or respondents of different ethnic/racial backgrounds; this difference in interpretation may have an impact in measures of self-reported health. 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 white 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 white women (Stevens, Kumanyika, and Keil 1994). Cultural differences in the meaning of the term “overweight” and attitudes about the acceptability of being overweight could well account for systematic response differences between African Americans and whites (Rajaram and Vinson 1998). True levels of a disorder among members of a certain group may be obscured if measures are used that do not take into account the cultural norms

of a particular group. Differences in meaning may be less of a problem with measures that do not rely on self-report; e.g., the body mass index can be calculated using physical measures of weight and height. However, for much health policy and epidemiological research, and for many constructs that affect health outcomes, such as depression and health beliefs, conceptual variations in self-reported measurement among different cultural groups may well impact substantially on measurement precision.

In the context of cross-cultural comparison, an important factor is consideration of the population of origin for instrument development, and whether the instrument has been tested for use with older adults in other populations, e.g., Mexican immigrants, or African Americans. Instruments that are not validated with respect to a particular racial or ethnic group are likely to carry different psychometric properties than is the one originally developed. For example, Fillenbaum et al. (1990) examined seven cognitive screening or neuropsychological tests in relation to clinical diagnosis. The authors reported that most measures, when adjusted for race and education, had lower specificities for African Americans than for whites. They suggested that most measures were culturally or educationally biased. Similarly, Teresi et al. (2001) reviewed studies of Differential Item Functioning (DIF) and item bias in the direct cognitive assessment measures with respect to race/ethnicity and education. Specifically, item performance varied across groups that differ in terms of education, ethnicity, and race (Teresi et al. 1995; Jones and Gallo 2002). Items that have shown high indices of validity and reliability for majority populations may lose their meaning when translated, as illustrated, for example, by the Spanish translation of the Mini-Mental State Exam (MMSE) (Folstein, Folstein, and McHugh 1975) item "no ifs, ands, or buts." This item has been found to be easier for Latinos than for non-Latinos (Valle et al. 1991; Teresi et al. 1995, 2001) possibly because of a translation artifact, i.e., the original intent of the item was lost in the translation. Thus, translations could alter the test properties, which in turn could result in a modification of those underlying abilities the test is measuring. 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 with those provided by clinical diagnosis. See Ramirez et al. (2001) for a review of the performance of cognitive screening measures across diverse populations in terms of sensitivity and specificity with respect to a clinical diagnosis.

When discussing such bias, particularly in the context of measurement accuracy, item structure and/or the criteria used in developing a measure

become highly relevant, as does the error introduced by the interviewer and/or the respondent (Teresi and Holmes 1997; Church 2001; van Hemert, Barendt, and Vermanda 2001). Contextually, raters and interviewers who come from different racial and ethnic backgrounds than do the individuals being rated/interviewed may respond to cues incorrectly, or in ways different from that intended, or may simply misinterpret information (Barisky 2000), leading to spurious study results (Shire 2002). For example, van Ryn and Burke (2000) in a study examining physicians' perceptions and beliefs and its potential implications for patients' diagnosis and treatment found that physicians (mainly white) were more likely to rate white patients as more educated and more rational than black patients even after controlling for patient's actual educational level. Although this finding can be simply explained by adherence to stereotypical beliefs that are inherently discriminatory, communication barriers such as differences in the patient's use of language when referring to symptoms, or symptom expression and/or interpretation of health-related behavior could possibly influence physicians' ratings across racial groups. As an example of an effort to address these issues, the Diagnostic and Statistical Manual of Mental Disorders—Fourth Edition (DSM—IV) Cultural Formulation provides guidelines regarding culturally sensitive statements that can be used in the assessment of symptoms and disorders (Neighbors et al. 2003).

Also problematic is the assumption of cultural homogeneity as it relates to measurement in ethnic populations that speak the same language. Cultural and idiomatic nuances can potentially exist within populations even though they share the same language. Examination of items such as “no ifs, ands, or buts” in eight (Lobo et al. 1979; Bird et al. 1987; Blesa et al. 1987; Tolosa, Alom, and Forcadell 1987; Gurland et al. 1992; Ortiz et al. 1997; Grupo de trabajo de neuropsicología 1999; Arias-Merino et al. 2003) independent Spanish versions of the MMSE used among Spanish-speaking populations inside and outside of the United States serve to illustrate the differences that can occur in item content and administration across such groups. The “no ifs, ands, or buts” item was different in all eight MMSE-Spanish versions; more importantly, it was different even among three independent versions developed in the same country, Spain (Lobo et al. 1979; Blesa et al. 1987; Tolosa, Alom, and Forcadell 1987). Similarly, two other versions (also different from each other) were used among study participants who were, most likely, Mexicans or of Mexican descent (one conducted in Guadalajara, Mexico; Arias-Merino et al. 2003 and the other in New Mexico, U.S.; Ortiz et al. 1997). Some of these adaptations of the item reflect an attempt to use seemingly linguistically equivalent expressions appropriate for the different Spanish-speaking

populations in question, while others attempted to represent the original intent of the item, i.e., measuring difficulties in the repeated articulation of consecutive consonants (see Ramírez et al. [under review], for a detailed discussion). Such differences may be relevant to study implementation, as well as to the comparisons and interpretations of findings. The presumption of social or cultural homogeneity exacerbates inaccurate cultural stereotypes, can lead to misleading conclusions in comparing prevalence of disorders, and can hinder the delivery of quality health care to different racial and ethnic groups.

Measurement error might lead to biased results (Smith and Reynolds 2002), and (in epidemiological research) to biased estimates of prevalence and of the magnitude of risk factors, and therefore for the development of public policies and service delivery (Skinner et al. 2001). Lack of fit between client needs and services rendered and/or public policies is the inevitable end result when cultural bias is introduced by the use of research instrumentation, which is insensitive to racial and ethnic differences.

In short, failure to account for inter- and intrarace variation creates problems for health care providers and/or program designers who often rely on research data as a basis for their decision making. Thus, there is a growing demand for the validation of existing measures using samples of minority group members, and for establishing the cross-ethnic equivalence of health-related assessment tools (Myers et al. 2000; Byrne and Watkins 2003). For example, the advisory Panel on Alzheimer's Disease (Advisory Panel on Alzheimer's Disease 1992) specifically calls for the development and validation of screening methods that will work effectively and fairly across various racial and ethnic groups.

MEASUREMENT ISSUES IN AGING AND MINORITY RESEARCH

Measurement equivalence in the context of cross-cultural research requires attention to both conceptual (or construct) and metric equivalence. Conceptual equivalence refers to whether or not constructs, domains, or behavior exemplars are the same or have the same meaning across compared groups, indicating that they are etic (generalizations that are "universally" valid) in nature. Translation artifacts (when scales are translated into languages different from the one in which they were originally developed), scale items using idiomatic expressions, terminology, and/or nomenclature that are relevant to

some racial and/or ethnic groups but not to others can result in conceptual nonequivalence.

The second interrelated concept is measurement or metric equivalence, which refers to whether or not the observed indicators relate to the latent factors in the same way across groups. Metric equivalence is assumed when similar factorial structures are found for the different racial and ethnic groups in question. Examination of factor loadings, measurement error variances, and factor means across groups can serve as indication of the degree of factorial invariance found across groups. Minimally, all measures marking the factors have to have their primary nonzero loadings on the same constructs across the multiple groups so that, arguably, factor scores can be compared in the context of cross-cultural research. This is sometimes referred to as configural invariance. However, most measurement experts argue that configural invariance is not sufficient and that metric invariance (meaning that the loadings on factors are equal across groups), is required to establish measurement equivalence.

Determination as to whether or not a measure is culturally fair cannot be made if metric equivalence is not first established. Some investigators discuss structural equivalence as a component of measurement equivalence. Structural equivalence is established when causal linkages among constructs and its causes and consequences are similar across compared groups. For instance, after it has been determined that a measure has conceptual and metric equivalence, determination has to be made as to whether or not its relationship with another measure is the same or different in terms of direction and/or magnitude in the different groups in question. However, an opposing view is that structural equivalence is not necessary, but relates to hypotheses to be tested regarding group differences in relationships among variables. The relationship among these concepts is hierarchical so that conceptual equivalence has to be established in order for measurement equivalence to be achieved. Therefore, addressing measurement comparability across groups that differ in culture or racial and ethnic background is usually a matter of the extent or the degree to which a specific measure shows comparability, determined by the type of equivalence that has been established (see, for example, Burnette 1998; Liang 2001; Mui, Burnette, and Chen 2001).

METHODS FOR IDENTIFYING BIAS

Within the research community, racial and ethnic measurement bias has been identified by some as a methodological issue requiring careful examination

(Teresi and Holmes 1994, 2001; Stewart and Napoles-Springer 2000). There are three broad approaches that have been used to assess the magnitude and nature of bias in measures when applied to diverse groups: qualitative studies, classic psychometric studies, and modern psychometric methods. Ideally, all three approaches should be applied sequentially, and/or iteratively during the development of measures.

Qualitative Studies

Qualitative studies can be used to assess the conceptual equivalence (and adequacy) of existing measures, e.g., to explore the relevance and appropriateness of concepts, and how individuals from diverse backgrounds give meaning to a particular domain (Gierl 2000; Stewart and Napoles-Springer 2000; Liang 2001). Qualitative studies also can help to determine whether any constructs are missing or interpreted differently across racial/ethnic groups. Qualitative approaches can also facilitate the understanding of how people construct their answers, e.g., the cognitive processes of reporting (Sundan, Bradburn, and Schwarz 1995) and to assess the level of congruence between the intent of the item and the respondent's interpretation of the question, e.g., random probe technique (Connidis 1983). Three commonly used qualitative methods in measurement studies are: cognitive testing (in-depth interviews, think-aloud interviews, behavioral coding interviews), focus groups, and expert panels.

Classical Psychometric Studies

Applications of traditional psychometric approaches have been used to examine measures across demographic subgroups. Procedures include an examination of content validity (a form of conceptual equivalence), and construct validity (including response bias and responsiveness to change). Examination of patterns of item variability and reliability, including interrater, test-retest, and internal consistency is performed. Finally, preliminary exploratory factor analysis and examination of dimensionality are often considered as part of the classical approach to scale development. Confirmatory factor analysis and tests of invariance are usually considered together with "modern psychometric theory" and latent variable approaches to measurement.

Modern Psychometric Methods: Latent Variable Approaches

Classical test theory-determined parameters and summary statistics are not invariant across groups given that they are reliant on the base rate of the

phenomenon being studied. For example, item variances, covariances, corrected item-total correlations, and α coefficients will vary from sample to sample and from subgroup to subgroup depending upon the prevalence of the condition being examined (Teresi and Holmes 1994). Thus, modern psychometric theory is being used increasingly to examine measurement properties, including metric equivalence, across demographic subgroups (Azocar et al. 2001; Teresi 2001; Fleishman, Spector, and Altman 2002). For example, about a decade ago the authors of an overview to the Annual Review of Gerontology and Geriatrics that focused on assessment summarized some of the statistical problems associated with the use of classical test theory methods for examination of bias in measures, and concluded that the use of modern psychometric theory, including item response theory (IRT) would become the standards for scale development in the twenty-first century (Teresi and Holmes 1994).

Ten years later this prediction has become fact. During the past 10 years the number of references to IRT in health-related research has risen from just a handful to several hundred. IRT has been used to develop new scales and to investigate old ones. IRT recently has been applied in the detection of DIF and item bias in epidemiological screening measures (Albert and Teresi 1999; Mungas et al. 2000; Teresi, Kleinman, and Welikson 2000; Teresi and Holmes 2001), and thus, for developing more accurate estimates of prevalence (Teresi et al. 1999). Applications of IRT such as computerized adaptive testing (CAT) are appearing in the health literature. CAT allows item selection to be targeted to individual disability level so that not all items or even the same items, need to be administered to everyone. However, such applications assume that the item bank from which items are selected has been examined for DIF (see Cook, O'Malley, and Roddey [2005] for a more detailed discussion of IRT and CAT).

As previously discussed, groups sharing a similar ethnic background, and even the same language may exhibit significant idiosyncratic and cultural differences (reflecting, for instance different acculturation levels) (Skinner 2001) that may need to be taken into account conceptually and methodologically. Such intragroup variations have received even less attention than intergroup differences in the context of psychometric testing. Application of a single model to all minority populations essentially ignores not only intergroup differences, but potential intragroup variations (e.g., within the Latino or within the African-American population); where the application of culturally sensitive parameters are, in like manner, necessary.

An example of metric equivalence, and of how one model may not fit all is provided by Gibson (1991) who, using latent variable confirmatory factor

analysis, examined racial differences in the structure and measurement of six self-reports of health widely used in studies involving older adults. The three elements of self-reported health examined were disease, disability, and subjective interpretation of health state. Findings showed that the form of the model had an overall acceptable fit for both the African-American and the white samples, indicating in this instance that for both groups, disease, disability, and subjective interpretations of health state derive from a single latent construct: internal perceptions of state of health. However, racial differences were seen in parts of the model, suggesting that culture and race affect the illness-reporting process in specific stages rather than as a whole (Gibson 1991). For example, subjective interpretation of health was a less valid measure of actual health state for African Americans than for whites, and the number of chronic conditions, as an indicator of disease, was a more valid measure for African Americans than for whites. As Gibson (1991) concluded, additional factors unique to each racial group that influence subjective interpretation of health state could be modeled. These differences in metric equivalence can influence structural relationships, resulting in erroneous conclusions about the underlying pathways in the disease process.

CONCLUSIONS

As research increasingly takes into account, and begins to focus on differences across diverse subgroups, issues of measurement comparability among these groups are paramount. Design and sampling issues must be considered carefully as they have bearings on the adequacy and generalizability of the compared population estimates. Furthermore, investigators performing comparative studies face the challenge of addressing measurement equivalence (see Liang 2001), crucial for obtaining accurate and substantive results in the context of cross-cultural comparisons. The argument is not that universally valid constructs or domains are not applicable cross-culturally, or that analyses focused on principles that are valid only within a given cultural system are the most appropriate, but that the assumption of universal applicability of standardized scales normed on particular cultural or racial/ethnic majority populations needs to be challenged and tested. In the context of cross-cultural research, the appropriateness of measures typically normed on cultural or racial/ethnic majority populations requires proper evaluation in order for the measures to be applicable to cultural or racial/ethnic minority groups so that relevant comparisons can be performed. To the extent that investigators

become acquainted with these issues, more research that examines the concepts and methodologies used in such studies will emerge. As a result, the adequacy of existing measures will be documented, the need for additional measures identified, and an agenda for future research developed.

ACKNOWLEDGMENTS

This material is based upon work supported by National Institutes of Health—Resource Centers for Minority Aging Research (RCAMR) at the Columbia Center for the Active Life of Minority Elders (P30 AG15294-06), at the Medical University of South Carolina (P30 AG21677) and at the University of California, San Francisco (P30 AG15272); and by the Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service, Measurement Excellence and Training Resource Information Center (METRIC; RES 02-235).

The views expressed herein are those of the authors and do not necessarily reflect those of the Department of Veterans Affairs.

REFERENCES

- Advisory Panel on Alzheimer's Disease. 1992. *Fourth Report of the Advisory Panel on Alzheimer's Disease*, NIH Publication No. 93-3520.
- Albert, S., and J. Teresi. 1999. "Reading Ability, Education and Cognitive Status Assessment among Older Adults in Harlem, New York City." *American Journal of Public Health* 89: 95–7.
- American Heart Association National Center. 2005. "Heart Facts 2003: Latino/Hispanic Americans" [accessed on March 14, 2005]. Available at <http://www.americanheart.org>
- Angel, R. J., and M. L. Frisco. 2001. "Self-Assessments of Health and Functional Capacity among Older Adults." *Journal of Mental Health and Aging* 7: 119–38.
- Arias-Merino, E. D., I. Orozco-Mares, L. C. Garabito-Esperanza, L. Fernandez-Cruz, M. J. Arias-Merino, A. Celis de la Rosa, C. Cabrera-Pivaral, and G. J. Gonzalez-Perez. 2003. "Correlates of Cognitive Impairment in Elderly Residents of Long Term Care Institutions in the Metropolitan Area of Guadalajara, Mexico." *Journal of Nutrition and Health in Aging* 7 (2): 97–101.
- Ashton, C. M., P. Haidet, D. A. Paterniti, T. C. Collins, H. S. Gordon, K. O'Malley, L. A. Petersen, B. F. Sharf, M. E. Suarez-Almazor, N. P. Wray, and R. L. Street. 2003. "Racial and Ethnic Disparities in the Use of Health Services." *Journal of General Internal Medicine* 18: 146–52.

- Azocar, F., P. Arean, J. Miranda, and R. F. Munoz. 2001. "Differential Item Functioning in a Spanish Translation of the Beck Depression Inventory." *Journal of Clinical Psychology* 57: 355-65.
- Barifsky, I. 2000. "The Role of Cognitive Equivalence in Studies of Health-Related Quality of Life Assessments." *Medical Care* 38: 125-9.
- Bird, H., G. Canino, M. Rubio-Stipec, and P. Shrout. 1987. "Use of Mini-Mental State Examination in a Probability Sample of Hispanic Population." *The Journal of Nervous and Mental Disease* 175 (12): 731-7.
- Blesa, R., M. Pujol, M. Aguilar, P. Santacruz, I. Bertran-Serra, G. Hernandez, J. M. Sol, and J. Peña-Casanova. 1987. "Clinical Validity of the 'Mini-Mental State' for Spanish Speaking Communities." *Neuropsychologia* 39: 1150-7.
- Bromberger, J. T., S. Harlow, N. Avis, H. M. Kravitz, and A. Cordal. 2004. "Racial/Ethnic Differences in the Prevalence of Depressive Symptoms among Middle-Aged Women: The Study of Women's Health across the Nation." *American Journal of Public Health* 94: 1378-85.
- Burnette, D. 1998. "Conceptual and Methodological Considerations in Research with Non-White Ethnic Elders." *Journal of Social Service Research* 23: 71-91.
- Byrne, B., and D. Watkins. 2003. "The Issue of Measurement Invariance Revisited." *Journal of Cross-Cultural Psychology* 34: 155-75.
- Cabassa, L. J. 2003. "Integrating Cross-Cultural Psychiatry into the Study of Mental Health Disparities." *American Journal of Public Health* 93: 1034-5.
- Church, A. T. 2001. "Personality Measurement in Cross-Cultural Perspective." *Journal of Personality* 69: 979-1006.
- Cohen, C. I., C. Magai, R. Yaffee, and L. Walcott-Brown. 2004. "Racial Differences in Paranoid Ideation and Psychoses in an Older Urban Population." *American Journal of Psychiatry* 161: 864-72.
- Connidis, I. 1983. "Integrating Qualitative and Quantitative Methods in Survey Research on Aging: An Assessment." *Qualitative Sociology* 6 (4): 334-53.
- Cook, K. F., K. J. O'Malley, and T. S. Roddey. 2005. "Dynamic Assessment of Health Outcomes: Time to Let the CAT Out of the Bag?" *Health Services Research* DOI: 10.1111/j.1475-6773.2005.00446.x.
- de Rekeneire, N., R. N. Rooks, E. M. Simonsick, R. I. Shorr, L. H. Kuller, A. V. Schwartz, and T. B. Harris. 2003. "Racial Differences in Glycemic Control in a Well-Functioning Older Diabetic Population: Findings from the Health, Aging and Body Composition Study." *Diabetes Care* 26: 1986-92.
- Federal Interagency Forum on Aging-Related Statistics. 2004. *Older American 2004: Key Indicators of Well-Being. Federal Interagency Forum on Aging Related Statistics*. Washington, DC: U.S. Government Printing Office.
- Fillenbaum, G., A. Heyman, K. Williams, B. Prosnitz, and B. Burchett. 1990. "Sensitivity and Specificity of Standardized Screens of Cognitive Impairment and Dementia among Elderly Black and White Community Residents." *Journal of Clinical Epidemiology* 43 (7): 651-60.
- Fiscella, K., P. Franks, M. R. Gold, and C. M. Clancy. 2000. "Inequality in Quality." *Journal of American Medical Association* 283: 2579-84.
- Fleishman, J. A., W. D. Spector, and B. M. Altman. 2002. "Impact of Differential Item Functioning on Age and Gender Differences in Functional Disability."

- Journal of Gerontology Series B: Psychological Sciences and Social Sciences* 57: S275–84.
- Folstein, M. F., S. E. Folstein, and P. R. McHugh. 1975. "The Mini-Mental State." *Journal of Psychiatric Research* 12: 189–98.
- Ford, M. E., and B. Hatchett. 2001. "Gerontological Social Work with Older African American Adults." *Journal of Gerontological Social Work* 36: 141–55.
- Frist, W. H. 2005. "Overcoming Disparities in U.S. Health Care." *Health Affairs* 24: 445–51.
- Gannotti, M. E., W. P. Handwerker, N. E. Groce, and C. Cruz. 2001. "Sociocultural Influences on Disability Status in Puerto Rican Children." *Physical Therapy* 81: 1512–24.
- Gibson, R. C. 1991. "Race and the Self-Reported Health of Elderly Persons." *Journal of Gerontology* 46: S235–42.
- Gierl, M. J. 2000. "Construct Equivalence on Translated Achievement Tests." *Canadian Journal of Education* 25 (2): 280–96.
- Grupo de trabajo de neuropsicología clínica de la Sociedad Neurológica Argentina. 1999. "El 'Mini-Mental State Examination': Instrucciones para su administración." *Revista Neurológica Argentina* 24 (1): 31–5.
- Guo, Y., D. B. Sigman, A. Borkowski, and N. Kyprianou. 2000. "Racial Differences in Prostate Cancer Growth: Apoptosis and Cell Proliferation in Caucasian and African-American Patients." *Prostate* 42: 130–6.
- Gurland, B., D. Wilder, P. Cross, J. Teresi, and V. Barrett. 1992. "Screening Scales for Dementia: Toward Reconciliation of Conflicting Cross-Cultural Findings." *International Journal of Geriatric Psychiatry* 7: 105–13.
- Jones, R. N., and J. J. Gallo. 2002. "Education and Sex Differences in the Mini-Mental State Examination: Effects of Differential Item Functioning." *Journal of Gerontology Series B—Psychological Sciences and Social Sciences* 57: P548–58.
- Kagawa-Singer, M., and N. Pourat. 2000. "Asian American and Pacific Islander Breast and Cervical Carcinoma Screening Rates and Healthy People 2000 Objectives." *Cancer* 89: 696–705.
- Kaplan, J. B., and T. Bennett. 2003. "Use of Race and Ethnicity in Biomedical Publication." *Journal of American Medical Association* 289: 2709–16.
- Kotchen, J. M., B. Shakoob-Abdullah, W. E. Walker, T. H. Chelius, R. G. Hoffmann, and T. A. Kotchen. 1998. "Hypertension Control and Access to Medical Care in the Inner City." *American Journal of Public Health* 88: 1696–9.
- Kressin, N. R., and L. A. Petersen. 2001. "Racial Differences in the Use of Invasive Cardiovascular Procedures: Review of the Literature and Prescription for Future Research." *Annals of Internal Medicine* 135: 352–66.
- Liang, J. 2001. "Assessing Cross-Cultural Comparability in Mental Health among Older Adults." *Journal of Mental Health and Aging* 7: 21–30.
- Lobo, A., J. Ezquerro, F. Gomez, J. M. Sala, and A. Seva. 1979. "'El Mini Examen Cognoscitivo' un test sencillo, práctico, para detectar alteraciones intelectivas en pacientes médicos." *Actas Luso-Españolas de Neurología y Psiquiatría* 3: 189–202.
- Mayberry, R. M., F. Mili, and E. Ofili. 2000. "Racial and Ethnic Differences in the Access of Medical Care." *Medical Care Research and Review* 57: 108–45.

- Moors, G. 2004. "Facts and Artifacts in the Comparison of Attitudes among Ethnic Minorities: A Multigroup Latent Class Structure Model Adjustment for Response Style Behavior." *European Sociological Review* 20: 303–20.
- Mui, A. C., D. Burnette, and L. M. Chen. 2001. "Cross-Cultural Assessment of Geriatric Depression: A Review of the CES-D and GDS." *Journal of Mental Health and Aging* 7: 137–64.
- Mungas, D., B. R. Reed, S. C. Marshall, and H. M. Gonzalez. 2000. "Development of Psychometrically Matched English and Spanish Language Neuropsychological Tests for Older Persons." *Neuropsychology* 14: 209–23.
- Mutran, E. J., P. S. Reed, and S. Sudha. 2001. "Social Support: Clarifying the Constructs with Applications for Minority Populations." *Journal of Mental Health and Aging* 7: 67–78.
- Myers, M. B., R. J. Catalone, T. J. Page Jr., and C. R. Taylor. 2000. "Academic Insights: An Application of Multiple-Group Causal Models in Assessing Cross-Cultural Measurement Equivalence." *Journal of International Marketing* 8: 108–22.
- Napoles-Springer, A., and A. Stewart. 2001. "Use of Health-Related Quality of Life Measures in Older Ethnically Diverse U.S. Populations." *Journal of Mental Health and Aging* 7: 173–80.
- National Institute of Diabetes and Digestive and Kidney Diseases. 2005. "National Diabetes Information Clearinghouse" [accessed on March 14, 2005]. Available at <http://www.niddk.nih.gov/health/diabetes/pubs/hispan/hispan.htm#t1>
- Neighbors, H. W., S. J. Trierweiler, B. C. Ford, and J. R. Muroff. 2003. "Racial Differences in DSM Diagnosis Using a Semi-Structured Instrument: The Importance of Clinical Judgment in Diagnosis of African Americans." *Journal of Health and Social Behavior* 44: 237–56.
- Ortiz, I. E., A. LaRue, L. J. Romero, M. F. Sassaman, and R. D. Lindeman. 1997. "Comparison of Cultural Bias in Two Cognitive Screening Instruments in Elderly Hispanic Patients in New Mexico." *American Journal of Geriatric Psychiatry* 5: 333–38.
- Powell, I. J., M. Banerjee, M. Novallo, W. Sakr, D. Grignon, D. P. Wood, and J. E. Pontes. 2000. "Prostate Cancer Biochemical Recurrence Stage for Stage Is More Frequent among African-American Than White Men with Locally Advanced but Not Organ-Confined Disease." *Urology* 55: 246–51.
- Rajaram, S. S., and V. Vinson. 1998. "African American Women and Diabetes: A Sociocultural Context." *Journal of Health Care for the Poor and Underserved* 9: 236–48.
- Ramírez, M., J. Teresi, D. Holmes, B. Gurland, and R. Lantigua. "Differential Item Functioning (DIF) and the Mini Mental Status Examination (MMSE): Overview Sample and Issues of Translation." *Medical Care* (under review).
- Ramírez, M., J. Teresi, S. Silver, D. Holmes, B. Gurland, and R. Lantigua. 2001. "Cognitive Assessment among Minority Elderly: Possible Test Bias." *Journal of Mental Health and Aging* 7: 91–118.
- Robin, R. W., R. L. Greene, B. Albaugh, A. Caldwell, and D. Goldman. 2003. "Use of the MMPI-2 in American Indians: Comparability of the MMPI-2 between Two Tribes and with the MMPI-2 Normative Group." *Psychological Assessment* 15: 351–9.

- Sankar, P., M. K. Cho, C. M. Condit, L. M. Hunt, B. Koenig, P. Marshall, S. S. J. Lee, and P. Spicer. 2004. "Genetic Research and Health Disparities." *Journal of American Medical Association* 291: 2985–9.
- Schneider, E. C., A. M. Zaslavsky, and A. M. Epstein. 2002. "Racial Disparities in the Quality of Care for Enrollees in Medicare Managed Care." *Journal of American Medical Association* 287: 1288–94.
- Scholderer, J., K. G. Grunert, and K. Brunso. 2005. "A Procedure for Eliminating Additive Bias from Cross-Cultural Survey Data." *Journal of Business Research* 58: 72–8.
- Shire, N. 2002. "Effects of Race, Ethnicity, Gender, Culture, Literacy, and Social Marketing on Public Health." *Journal of Gender Specific Medicine* 5: 48–54.
- Sinclair, S., P. Hayes-Reams, H. F. Myers, W. Allen, J. Hawes-Dawson, and R. King-ton. 2002. "Recruiting African-Americans for Health Studies: Lessons from the Drew-RAND Center on Health and Aging." *Journal of Mental Health and Aging* 6: 39–51.
- Skinner, J. 2001. "Acculturation: Measures of Ethnic Accommodation to the Dominant American Culture." *Journal of Mental Health and Aging* 7: 41–52.
- Skinner, J. H., J. A. Teresi, D. Holmes, S. M. Stalh, and A. L. Stewart. 2001. "Measurement in Older Ethnically Diverse Populations: Overview of the Volume." *Journal of Mental Health and Aging* 7: 5–11.
- Smedly, B. D., A. Y. Stith, and A. R. Nelson, (Eds) 2002. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: Institute of Medicine.
- Smith, A. M., and N. L. Reynolds. 2002. "Measuring Cross-Cultural Service Quality: A Framework for Assessment." *International Marketing Review* 19: 450–82.
- Snowden, L. 2003. "Racial/Ethnic Bias and Health Bias in Mental Health Assessment and Intervention: Theory and Evidence." *American Journal of Public Health* 93: 239–43.
- Stanley, S., and J. Chang. 2003. "The State of Psychological Assessment in Asia." *Psychological Assessment* 15: 306–10.
- Stevens, J., S. K. Kumanyika, and J. E. Keil. 1994. "Attitudes toward Body Size and Dieting: Differences between Elderly Black and White Women." *American Journal of Public Health* 84: 1322–5.
- Stewart, A. L., and A. Napoles-Springer. 2000. "Health-Related Quality of Life Assessments in Diverse Population Groups in the United States." *Medical Care* 38: II125–9.
- Sundan, S., N. M. Bradburn, and N. Schwarz. 1995. *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass Inc.
- Tanjasiri, S. P., and L. Sablan-Santos. 2001. "Breast Cancer Screening among Chamorro Women in Southern California." *Journal of Women's Health and Gender Based Medicine* 10: 479–85.
- Teresi, J., S. Albert, D. Holmes, and R. Mayeux. 1999. "Use of Latent Class Analyses for 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* 18: 309–21.

- Teresi, J., D. Holmes, M. Ramirez, B. Gurland, and R. Lantigua. 2001. "Performance of Cognitive Tests among Different Racial/Ethnic and Education Groups: Findings of Differential Item Functioning and Possible Test Bias." *Journal of Mental Health and Aging* 7: 79–89.
- Teresi, J., M. Kleinman, and K. Welikson. 2000. "Modern Psychometric Methods for Detection of Differential Item Functioning: Application to Cognitive Assessment Measures." *Statistics in Medicine* 19: 1651–83.
- Teresi, J. A. 2001. "Statistical Methods for Examination of Differential Item Functioning (DIF) with Applications to Cross-Cultural Measurement of Functional, Physical and Mental Health." *Journal of Mental Health and Aging* 7: 31–40.
- Teresi, J. A., R. R. Golden, P. Cross, B. Gurland, M. Kleinman, and D. Wilder. 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: 473–83.
- Teresi, J. A., and D. Holmes. 1994. "Overview of Methodological Issues in Gerontological and Geriatric Measurement." In *Annual Review of Gerontology and Geriatrics: Focus on Assessment Techniques*, edited by M. P. Lawton and J. Teresi, pp 1–23. New York: Springer Publishing Company.
- . 1997. "Methodological Issues in Cognitive Assessment and Their Impact on Outcomes Measurement." *Alzheimer Disease and Associated Disorders* 11: 146–55.
- . 2001. "Some Methodological Guidelines for Cross-Cultural Comparisons." *Journal of Mental Health and Aging* 7: 13–9.
- Tolosa, E., J. Alom, and F. Forcadell. 1987. "Criterios diagnósticos y escalas evaluativas en la enfermedad de Alzheimer." *Revista Clínica Española* 181: S56–9.
- Tranh, T. V., D. Ngo, and K. Conway. 2003. "A Cross-Cultural Measure of Depressive Symptoms among Vietnamese Americans." *Social Work Research* 27: 56–65.
- Turner, R. J., and W. R. Avison. 2003. "Status Variation in Stress Exposure: Implications for the Interpretation of Research on Race, Socioeconomic Status, and Gender." *Journal of Health and Social Behavior* 44: 488–506.
- Valle, R., R. Hough, B. Kolody, H. Cook-Gait, G. F. Velazquez, and R. Jimenez. 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.
- van Hemert, D. A., C. Baerveldt, and M. Vermande. 2001. "Assessing Cross-Cultural Item Bias in Questionnaires: Acculturation and the Measurement of Social Support and Family Cohesion for Adolescents." *Journal of Cross-Cultural Psychology* 4: 381–97.
- van Ryn, M., and J. Burke. 2000. "The Effect of Patient Race and Socio-Economic Status on Physician's Perceptions of Patients." *Social Science and Medicine* 50: 813–28.
- Villa, V. M., and M. P. Aranda. 2000. "The Demographic, Economic, and Health Profile of Older Latinos: Implications for Health and Long-Term Care Policy

- and the Latino Family.” *Journal of Health and Human Services Administration* 23: 161–80.
- Young, B. A., C. Maynard, and E. J. Boyko. 2003. “Racial Differences in Diabetic Nephropathy, Cardiovascular Disease, and Mortality in a National Population of Veterans.” *Diabetes Care* 26: 2392–9.
- Zingmond, D. S., A. M. Kilbourne, A. C. Justice, N. S. Wenger, M. Rodriguez-Baradas, L. Rabeneck, D. Taub, S. Weissman, J. Briggs, J. Wagner, S. Smola, and S. A. Bozzette. 2003. “Differences in Symptom Expression in Older HIV-Positive Patients: The Veterans Aging Cohort 3 Site Study and HIV Cost and Service Utilization Study Experience.” *Journal of Acquired Immune Deficiency Syndromes* 33: S84–92.