

# Using the NIH Toolbox in special populations

Considerations for assessment of pediatric, geriatric, culturally diverse, non-English-speaking, and disabled individuals

David Victorson, PhD  
Jennifer Manly, PhD  
Kathleen Wallner-Allen,  
PhD  
Nathan Fox, PhD  
Christy Purnell, BA  
Hugh Hendrie, MB,  
ChB, DSc  
Richard Havlik, MD  
Mark Harniss, PhD  
Susan Magasi, PhD  
Helena Correia  
Richard Gershon, PhD

Correspondence to  
Dr. Victorson:  
d-victorson@northwestern.edu

## ABSTRACT

**Background:** In order to develop health outcomes measures that are relevant and applicable to the general population, it is essential to consider the needs and requirements of special subgroups, such as the young, elderly, disabled, and people of different ethnic and cultural backgrounds, within that population.

**Methods:** The NIH Toolbox project convened several working groups to address assessment issues for the following subgroups: pediatric, geriatric, cultural, non-English-speaking, and disabled. Each group reviewed all NIH Toolbox instruments in their entirety.

**Results:** Each working group provided recommendations to the scientific study teams regarding instrument content, presentation, and administration. When feasible and appropriate, instruments and administration procedures have been modified in accordance with these recommendations.

**Conclusion:** Health outcome measurement can benefit from expert input regarding assessment considerations for special subgroups. *Neurology*<sup>®</sup> 2013;80 (Suppl 3):S13-S19

## GLOSSARY

**AWG** = Accessibility Working Group; **CWG** = Cultural Working Group; **GWG** = Geriatric Working Group; **PWG** = Pediatric Working Group.

There is increasing acknowledgment that certain subgroups of the US population are more vulnerable to receiving inequitable health care, and thus achieving inferior health outcomes compared with other groups.<sup>1</sup> Such vulnerability may relate to one's age, race, ethnicity, primary language, economic status, sex, and ability/disability status.<sup>2,3</sup> A cornerstone of equitable health care begins with well-developed health outcomes assessment tools. It is essential that the design, development, and implementation of an assessment tool, including the interpretation of subsequent findings, consider the challenges, opportunities, and special circumstances of the vulnerable subgroups mentioned previously. The purpose of this article is to highlight progress that has been made among several working groups for the NIH Toolbox for the Assessment of Neurological Behavior and Function (NIH Toolbox)<sup>4</sup> that were formed to address the needs and concerns of health outcomes assessment in the following populations: pediatric, geriatric, cultural, non-English-speaking, and disability.

Working groups comprised national and international experts in their respective fields. Although groups may have varied in specific methodologies used to accomplish their goals, each group reviewed all NIH Toolbox measures in their entirety and provided recommendations to the scientific study team. These recommendations were considered for inclusion in the NIH Toolbox and supplemental instruments with available time and practicality being a major consideration. The following sections provide an overview of work that was accomplished and feature some of the major issues that were encountered as well as strategies for remediation. Complete working group reports can be found at the NIH Toolbox Web site: [www.nihtoolbox.org](http://www.nihtoolbox.org).

From the Northwestern University Feinberg School of Medicine (D.V., S.M., H.C., R.G.), Chicago, IL; Columbia University (J.M.), New York, NY; Westat, Inc. (K.W.-A., R.H.), Rockville, MD; University of Maryland (N.F.), College Park, MD; Indiana University School of Medicine (C.P., H.H.), Indianapolis, IN; and University of Washington (M.H.), Seattle, WA.

Go to [Neurology.org](http://Neurology.org) for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

## CHALLENGES AND OPPORTUNITIES FOR PEDIATRIC ASSESSMENT

**Kathleen Wallner-Allen and Nathan Fox.** Conducting research with children poses several challenges, many of which are magnified in longitudinal research because there are few measures that can be used across different age points, making it more difficult to measure or understand change over time.<sup>5,6</sup> Pediatric assessment is also difficult because performance differences are often attributed to factors other than the construct of interest (e.g., complexity of instructions, testing environment). A Pediatric Working Group (PWG) was created to help address the many challenges to creating measures appropriate across a broad age span and to help focus on design and procedural issues of importance for young children through adolescents. The PWG comprised scientists who have extensive experience conducting research with children and who have expertise across a broad range of issues in child development and with large-scale data collection. Its purpose was to evaluate the extent to which the NIH Toolbox measures were appropriate across the 3- to 18-year age range, to raise issues relevant to this age group that may not have been considered, and to make specific recommendations for how to improve instruments and procedures for children. Over the past 2 years, the PWG has consulted the extant literature and has interacted through conference calls, WebEx online meetings, and extensive e-mail discussions, with members having hands-on opportunities with the proposed NIH Toolbox instruments.

During the review of potential instruments, potential threats to validity were noted and suggestions for improvement were made. By providing simple, clear, easy-to-follow instructions, using task materials that are engaging, concrete, and familiar, and structuring the testing environment in a child-friendly way, differences in task performance across different ages will more likely reflect differences in competence on the construct of interest rather than differences in performance factors such as the ability to follow complex instructions. The review indicated that all domains needed to reduce the complexity of the language used for instructions, in the way tasks were explained, and/or in questionnaire items. Domain teams were also encouraged to provide task-specific guidelines to standardize the interactions between the examiner and participant. When, what, and how often the examiner can say something during task administration is of particular importance for young children.

An early challenge for the PWG was to provide guidance on how to test children across the 4 domains of the NIH Toolbox. Children require special considerations for testing including instructions presented slowly and simply (and orally to children who are not yet able to read), training trials to ensure comprehension, attractive stimuli to keep a child's attention, and response hardware that is appropriate for small

hands. As a result of these considerations, a set of pediatric assessment principles were articulated. They addressed such issues as instrument design characteristics, the testing environment, the psychological and physical needs of the child, and the nature and extent of the interactions among the test administrator, the child or adolescent, and the parent.

The group made design recommendations for the computer interface for NIH Toolbox instruments, advising on the use of a touch screen, a mouse, or an alternative response mechanism and on whether instructions should be "live" or prerecorded and provided over the computer to standardize the presentation of task instructions. Recommendations were also made for the quality of the voice and the sex of the speaker, and experts listened to voice samples to ensure that the recommendations were captured by the voice selected to record task instructions. Pediatric experts also suggested that instruments have built-in flexibility with the computer interface to allow for the possibility of repeating instructions. Emphasis was placed on ensuring that the language used in instructions was not too complex regarding individual words, the complexity of the sentence structure, and/or the complexity with which the task was explained.

## GERIATRIC ASSESSMENTS

**Christy Purnell, Hugh Hendrie, and Richard Havlik.** Assessment in older individuals has to take into consideration a number of issues. These include the changes associated with aging; for example, relative psychomotor slowing of response, the accumulation of physical and sensory impairments in older adults, and the possible unfamiliarity with recent technology, including the use of computers in this population.<sup>7-9</sup> To ensure that the NIH Toolbox instruments were suitable for use in older adults, a Geriatric Working Group (GWG) was formed. Meeting regularly through a series of conference calls, members of the GWG reviewed the relevant literature, and each committee member was encouraged to communicate with local experts in test administration to older adults, particularly those tests that were computer based.

As part of this process, the GWG developed a document outlining the principles of geriatric assessment (available at [www.nihtoolbox.org](http://www.nihtoolbox.org)), which contains 3 sections. Section I addressed general issues related to assessment of older participants and was based to a large extent on a National Institute on Aging report "Talking with Your Older Patient."<sup>10</sup> This document was most pertinent to issues relating to test administration. It included recommendations such as establishing respect, avoiding hurrying, using understandable terms, assuring comfort, being alert for sensory impairments, being alert for signs of stress, fearfulness, fatigue, and physical distress, and considering that

performance may be affected by medication use or time of the day of administration. Section II, the concept of universal design, was based on a document supported by the National Institute on Disability and Rehabilitation Research.<sup>11</sup> It addressed the principle that measures should be developed for use by as many people as possible. This document included information on equitable use for individuals with diverse abilities, flexibility in use to accommodate a wide range of individual abilities, simple and intuitive use, effective communication, low physical effort, and size and space for approach and use. Section III addressed the issue of computer use in older adults. It included a review of the growing literature on technology use for older adults as illustrated by the new official publication of the International Society for Gerontechnology, *The Journal of Gerontechnology*. It was also based on advice received from local experts in the field and included considerations of issues related to a) computer presentation (for example, the use of a large screen or large text sizes), and b) input, such as the use of a trackball rather than a mouse and instrument adaptations for individuals with upper extremity motor impairments.

Using these principles, the GWG then reviewed each of the instruments developed by the different domain teams. The GWG determined that, because many of the problems relating to instrument suitability for older adults were associated with physical or sensory impairment, it would work closely with the NIH Toolbox Accessibility Committee in preparing recommendation reports. The resulting recommendations addressed the following areas:

1. *Test administration.* The necessity for appropriate training of a test administrator to accommodate older participants was emphasized. In the cognitive tests, for example, it was believed that the instructions may be too complicated if an administrator was not present. The administration should also ensure the comfort and safety of the participants. A specific example of the latter was the recommendation to use a gait belt to prevent falls in the NIH Toolbox Standing Balance Tests.
2. *Production and presentation.* Production and presentation of the test instruments were particularly relevant for instruments that were computer based. This included consideration of such issues as icon size, font size, number of items on the screen, complexity of instructions, the pros and cons regarding the use of a touch screen, and availability of alternative response options. Examples included the adoption of a less-confusing introductory segment in the Imitation Based Assessment of Memory and the use of similar picture axes in the List Sorting test.

3. *Item composition.* These included the wording of the items, their understandability, and their suitability for use in an older population. For example, in the emotional domain, there was concern about the use of different time frames (i.e., 7 days, 2 weeks, 30 days) for contiguous questions and response variables that were too similar (i.e., not at all, not at all sure, not at all true).

Recommendations from the GWG were presented to each domain leader and suitable modifications to the instruments were made, resulting in instruments that are more user friendly for older adults.

### **CULTURAL CONSIDERATIONS** David Victorson,

Jennifer Manly, and Helena Correia. Given the increasing cultural diversification of the US population, it becomes imperative to evaluate and ensure the cultural competency of health outcomes measurement tools with the same fervor that is used to assess a tool's psychometric characteristics.<sup>12-14</sup> As such, an internationally recognized Cultural Working Group (CWG) was assembled to evaluate the extent to which NIH Toolbox measures are culturally sensitive and conceptually appropriate across different cultural groups and to make specific recommendations to the NIH Toolbox scientific study team that highlight their strengths, limitations, and strategies for remediation.

Through a series of e-mail and teleconference communications, as well as a face-to-face meeting, the CWG first consulted the extant literature to establish cultural review standards by which NIH Toolbox measures should be evaluated.<sup>12,15-18</sup> These efforts led to the following 5 criteria considered by the CWG to be imperative in ensuring cultural competency:

1. The perspectives of culturally diverse individuals must be incorporated into the NIH Toolbox's development.
2. An equality of conceptual, semantic, and/or linguistic meaning across different cultural groups must be achieved.
3. A metric must be created to evaluate the extent that NIH Toolbox measures are fair across cultural groups (e.g., face validity), reflective of their intended construct (e.g., content validity), culturally biased/inappropriate/offensive, unlikely to be misinterpreted (e.g., measurement artifacts), considered alongside relevant socioeconomic data, and are subject to analysis that examines procedural equivalence (through use of confirmatory factor analysis, item response theory, differential item functioning), and culturally representative in their sampling plans for large-scale testing.
4. Procedures must be identified to address existing differential item functioning across racial/ethnic groups.

5. Technical equivalence must be ensured/measurable (e.g., whether different cultural or lower socioeconomic status groups respond equally to technical measurement properties, such as Likert-type scales).

After setting the aforementioned criteria, members of the CWG convened for a day-long face-to-face meeting to review NIH Toolbox measurement tools within the context of these cultural competency criteria. This included an examination of both the comprehensive translation methodology that was used and the proposed sociodemographic form, as well as an in-depth review of each of the NIH Toolbox instruments in cognition, sensory, motor, and emotion domains. Each measure was evaluated either through actual administration or through video presentations. In addition to domain-specific recommendations, several suggestions were provided to ensure greater knowledge of the cultural and economic diversity of the sample, including gathering information on race/ethnicity, highest level of education completed, education quality, location of education, acculturation, immigration experience, and nationality. Recommendations were considered regarding the NIH Toolbox norming phase. Of importance was the topic of bilingual respondents and the need to inquire about the extent of formal education completed in each language spoken. Additional discussion centered on how to best capture racial/ethnic self-identification, nation and state of birth, immigration and acculturation experience, and parental country of origin. For the language spoken at home, members recommended that children be asked what they speak with their caregivers and friends, and that adults be asked what language is spoken at home and with partners. Several members of the CWG believed that income does not sufficiently capture socioeconomic status and that a better measure could be attained if income was combined with additional data on the participant's state and city of birth, current zip code (attainable for both United States and foreign-born, and via the US Census for urban/rural based on zip code), and whether the person was born in an urban or rural area.

### TRANSLATION AND LINGUISTIC CONSIDERATIONS

**Helena Correia.** Creating a Spanish version of the NIH Toolbox enables assessment of Spanish speakers who do not speak English or do not speak it well. However, given the heterogeneity of the Spanish-speaking population in the United States, it is particularly challenging to develop a Spanish version that is suitable for all Spanish-speaking individuals. Depending on country of origin, the names given to certain mundane objects or situations can differ. “Banana,” for example, can be plátano, banana, banano, cambur, guineo, inguiri, and

platanito depending on to whom one talks. The issue of universal wording is especially significant in the context of memory-based and image identification tasks, because they involve word retrieval. Relying on a computer-assisted evaluation with an audio component, some of the NIH Toolbox measures pose additional challenges for translation. Issues such as sex and form of address must be dealt with in a manner suitable for written as well as oral form. Sex affects primarily self-report items, because adjectives have different endings depending on the sex of the person who is self-reporting. The sex-inclusive option used in written form (e.g., “deprimido/a” for “depressed”) is not feasible for audio delivery of the item, however, where 2 sex-specific versions would be necessary. Form of address comes into play when translating instructions because verbs and pronouns have different forms in Spanish depending on whether the recipient of the instruction is a child or an adult (i.e., informal vs formal form of address). For example, an English instruction can say “choose the answer that shows how you feel” to address both adults and children, but when translated into Spanish, the same instruction requires 2 versions.

Translating instruments that were developed in English for a non-English-speaking audience can be challenging because often the wording or the concepts have no equivalent in other languages. Using clear syntax, simple and common language, and avoiding extreme colloquialisms contribute to improving comprehension of the English source and to a certain extent that of the translated version. However, that is not always enough to enable equivalent translations. Spanish does not have, for example, a term to encompass all the negative emotions contained in the multidimensional but simple English concept of “upset.” In order to translate it, it is necessary to clarify its intended meaning in the particular item. Consideration of language register also affects the creation of the NIH Toolbox Spanish version in that for some measures, the criteria applied to the development of the English version are not equally applicable to the Spanish version. For example, the words from the Vocabulary list, if simply translated, do not reflect the same level of difficulty in both languages (e.g., “moribund” belongs to a higher register in English whereas “moribundo” is easier to understand in Spanish).

A Spanish Language Working Group was formed to identify specific problems through a translatability review during instrument development and to offer alternative wording solutions more suitable for a culturally diverse population, for translation, and for the survey's mode of administration. Next, a comprehensive translation methodology was used to produce a universal Spanish version appropriate for the majority of US Spanish-speakers. During the translatability review of individual

NIH Toolbox measures, the primary focus was on the “word-heavy” instruments from Emotion and Cognition. However, selected measures from Sensory were also assessed. For Motor, an early translation of 2 of the measures revealed no significant translatability issues. Although different approaches were adopted for each of the domains based on the specificity of the measures, the following criteria were used to assess translatability: universality, cultural relevance, figure of speech/jargon, ambiguity, register, number of words, translation reversal, double-negative, double-barrel, sex and number agreement, parts of speech, oral vs written, and mode of administration and technology.

All self-report items were translated according to the Functional Assessment of Chronic Illness Therapy translation methodology,<sup>19–22</sup> which involves 1) 2 simultaneous forward translations; 2) reconciled single Spanish translation conducted by a third independent translator; 3) back-translation by a native English-speaking translator; 4) comparison of source and back-translated versions to identify discrepancies; 5) reviews from 3 bilingual experts; 6) finalization by the Spanish language coordinator; 7) harmonization and quality assurance; 8) formatting, typesetting, and proofreading; and 9) cognitive pretesting via interviews with participants who are native speakers of Spanish.

**DISABILITY ACCESS AND THE NIH TOOLBOX** Mark Harniss and Susan Magasi. In the United States, approximately 15% of community-dwelling individuals older than 5 years have a disability, and the prevalence of disabling conditions increases with age.<sup>23,24</sup> Federal law, such as the Americans with Disabilities Act, mandates that these 41.3 million people with disabilities have the right to full participation and equal access to all aspects of society. As such, end users of the NIH Toolbox (e.g., NIH grantees) will encompass people with disabilities as well as people with sensory, motor, and cognitive impairments in their research.

To ensure that the NIH Toolbox measures are accessible to persons with disabilities, an Accessibility Working Group (AWG) was formed composed of individuals who have experience with disability, accessible information technology, and assistive technologies. A first task of the AWG was to develop guidelines based on Section 508 of the Rehabilitation Act and the Web Consortium Accessibility Guidelines and disseminate them to the motor, cognition, emotion, and sensory project teams. Next, the AWG conducted a review of each proposed measure and identified whether it was a) accessible, b) inaccessible but could potentially be made accessible, or c) inaccessible and likely could not be made accessible. In the latter category, they also differentiated measures that were not accessible but should be and measures that were

not accessible and should not be. For example, a measure of working memory probably should be accessible to people who are blind, but a measure of visual acuity did not need to be altered for people who were blind because it was a measure of vision. The AWG considered a broad range of functional limitations such as vision, hearing, motor, speech, and reading. Measures were reviewed regardless of whether they involved information technology (i.e., fell under Section 508 guidelines) or not. This included a detailed task analysis to identify the motor, cognitive, and sensory demands of each measure.

In their original form, all of the NIH Toolbox measures presented accessibility challenges for different groups of people. Individuals with vision impairments (blind or low vision) face the greatest accessibility challenges, and individuals with motor impairments are a close second. Accessibility challenges fall into 3 broad categories: 1) some measures are inaccessible because a construct is being measured in a way that requires a specific functional ability (e.g., a measure of vocabulary that requires vision or a measure of endurance that requires walking); 2) some measures are inaccessible because administration guidelines are narrowly defined (e.g., a measure of taste that gives directions only verbally); 3) some measures are inaccessible because of decisions made about technology (e.g., a pattern-comparison measure that requires users to make use of a touch screen—an instrument since modified by the instrument developers).

The AWG reported its findings to the project teams and offered technical assistance to assessment and technology developers on accessibility problems and potential solutions. Recommendations to improve the accessibility of all candidate measures in the NIH Toolbox challenged instrument developers to re-evaluate testing procedures and disentangle mode of administration from the construct under evaluation. Key recommendations included ensuring redundancy in mode of presentation of both the instrument content and response options. This includes adjusting administration guidelines; for example, acknowledging that sign language and written directions are acceptable and re-evaluating the technology platform and input options to determine the most accessible software and hardware implementations. Some developers have clarified the construct they are evaluating (e.g., the episodic memory task actually only focuses on visual episodic memory and, unfortunately, therefore is not applicable for the visually impaired). Had the construct been verbal memory, the task would have not been applicable for the hearing impaired. Next steps involve identifying changes that are feasible and necessary to make vs changes that need to be made but are not feasible as part of the NIH Toolbox development. Evaluation and modification of NIH Toolbox instruments' accessibility can lead to the development of universally

accessible health outcomes measurement and research for millions of Americans with disabilities.

### AUTHOR CONTRIBUTIONS

David Victorson: drafting/revising the manuscript, study concept or design, analysis or interpretation of data, acquisition of data, study supervision. Jennifer Manly: drafting/revising the manuscript, contribution of vital reagents/tools/patients. Kathleen Wallner-Allen: drafting/revising the manuscript, study concept or design, analysis or interpretation of data. Nathan Fox: drafting/revising the manuscript, study concept or design, analysis or interpretation of data, study supervision. Christy Purnell: analysis or interpretation of data, acquisition of data. Hugh Hendrie: drafting/revising the manuscript, study concept or design, analysis or interpretation of data, acquisition of data, study supervision. Richard Havlik: study concept or design, study supervision. Mark Harniss, drafting/revising the manuscript, study concept or design, analysis or interpretation of data, acquisition of data, study supervision. Susan Magasi: drafting/revising the manuscript, study concept or design, acquisition of data, study supervision. Helena Correia: drafting/revising the manuscript, coordinating all translation-related activities and verifying quality of the translation. Richard Gershon: drafting/revising the manuscript, contribution of vital reagents/tools/patients, acquisition of data, statistical analysis, obtaining funding.

### STUDY FUNDING

This project is funded in whole or in part with Federal funds from the Blueprint for Neuroscience Research and the Office of Behavioral and Social Sciences Research, NIH, under contract no. HHS-N-260-2006-00007-C.

### DISCLOSURE

D. Victorson holds stock options in Eli Lilly and Company, received an honoraria for serving on the Steering Committee of the Reeve Neuro-Recovery Network, was funded by NIH contracts HHSN265200423601C and HHS-N-260-2006-00007-C and grants R01HD054569-02NIDRR, 1U01NS056975-01, and R01 CA104883, received support from the American Cancer Society (National and Illinois Division) for research in prostate cancer, received institutional support from North-Shore University HealthCare System for research in prostate cancer, received institutional support from the Medical University of South Carolina for sarcoidosis research, and received institutional support from the Northwestern Medical Faculty Foundation for urology research. J. Manly is funded by NIH grants R01AG028786 and R01AG037212; she had received funding previously from NIH grant R01AG016206 and a grant from the Alzheimer's Association (IIRG 05-14236). K. Wallner-Allen reports no disclosures. N. Fox is funded by NIH grants R37HD017899, MH074454, U01MH080759, R01MH091363, P50MH078105, and P01HD064653. He serves on the scientific board of the National Scientific Council for the Developing Child. C. Purnell reports no disclosures. H. Hendrie currently receives research funding from NIH/NIA grant(s) R01AG009956, R24MH080827, 5R01AG026096-05, UF20303/U01AG022376, R01AG031222, R01AG019181, and R01AG029884. R. Havlik reports no disclosures. M. Harniss is funded by multiple grants and contracts from the NIH (HHS-N-260-2006-00007-C), the National Institute of Disability and Rehabilitation Research (H133B080025, H133B080024, and H133A110014), and the Elections Assistance Commission. S. Magasi is funded by multiple grants and contracts from the NIH (HHS-N-260-2006-00007-C, 1U5AR057951-01, and U01 AR 052177), the National Institute of Disability and Rehabilitation Research (H133B090024), Agency on Healthcare Research and Quality (RFA-HS-11-001), with minor funding from Forest-Ironwood Pharmaceuticals, and Dai-ichi-Sankyo, Inc. H. Correia reports no disclosures. R. Gershon has received personal compensation for activities as a speaker and consultant with Sylvan Learning, Rockman, and the American Board of Podiatric Surgery. He has several grants awarded by NIH: N01-AG-6-0007, 1U5AR057943-01, HHSN260200600007, 1U01DK082342-01, AG-260-06-01, HD05469; NINDS: U01 NS 056 975 02; NHLBI K23: K23HL085766; NIA: 1RC2AG036498-01; NIDRR: H133B090024; OppNet: N01-AG-6-0007. Go to [Neurology.org](http://Neurology.org) for full disclosures.

Received June 6, 2012. Accepted in final form August 1, 2012.

### REFERENCES

1. Hahn EA, Cella D. Health outcomes assessment in vulnerable populations: measurement challenges and recommendations. *Arch Phys Med Rehabil* 2003;84:S35–S42.
2. Quality First: Better Health Care for All Americans. Final Report of the President's Advisory Commission on Consumer Protection and Quality in the Health Care Industry. Washington, DC: US Government Printing Office; 1998.
3. United States Agency for Healthcare Research Quality. National healthcare disparities report 2010 [online]. Available at: <http://www.ahrq.gov/qual/nhdr10/nhdr10.pdf>. Accessed February 4, 2013.
4. Gershon RC, Cella D, Fox NA, Havlik RJ, Hendrie HC, Wagster MV. Assessment of neurological and behavioural function: the NIH Toolbox. *Lancet Neurol* 2010;9:138–139.
5. Kozinetz CA, Warren RW, Berseth CL, Aday LA, Sachdeva R, Kirkland RT. Health status of children with special health care needs: measurement issues and instruments. *Clin Pediatr* 1999;38:525–533.
6. Knox SS, Echeveria D. Methodological issues related to longitudinal epidemiological assessment of developmental trajectories in children. *J Epidemiol Community Health* 2009;63:1–3.
7. Maly RC, Hirsch SH, Reuben DB. The performance of simple instruments in detecting geriatric conditions and selecting community-dwelling older people for geriatric assessment. *Age Ageing* 1997;26:223–231.
8. McHorney CA. Ten recommendations for advancing patient-centered outcomes measurement for older persons. *Ann Intern Med* 2003;139:403–409.
9. Naeim A, Reuben D. Geriatric syndromes and assessment in older cancer patients. *Oncology* 2001;15:1567–1577, 1580.
10. National Institute on Aging. Talking with Your Older Patient: A Clinician's Handbook. Bethesda: US Department of Health and Human Services, National Institutes of Health, National Institute on Aging; 2008.
11. Connell B, Jones M, Mace R. Principles of Universal Design. Raleigh: North Carolina State University, Center for Universal Design; 1997.
12. Manly JJ. Deconstructing race and ethnicity: implications for measurement of health outcomes. *Med Care* 2006;44:S10–S16.
13. Prince M. Measurement validity in cross-cultural comparative research. *Epidemiol Psychiatr Soc* 2008;17:211–220.
14. Strickland OL. Cultural considerations and issues in measurement. *J Nurs Meas* 2003;11:3–4.
15. Bullinger M, Anderson R, Cella D, Aaronson N. Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. *Qual Life Res* 1993;2: 451–459.
16. Dana RH. Culturally competent assessment practice in the United States. *J Pers Assess* 1996;66:472–487.
17. Teresi JA. Statistical methods for examination of differential item functioning (DIF) with applications to cross-cultural measurement of functional, physical and mental health. *J Ment Health Aging* 2001;7:31–40.
18. American Psychological Association. Guidelines on multicultural education, training, research, practice, and organizational change for psychologists—American Psychological Association. *Am Psychol* 2003;58:377–402.
19. Bonomi AE, Cella DF, Hahn EA, et al. Multilingual translation of the Functional Assessment of Cancer Therapy (FACT) quality of life measurement system. *Qual Life Res* 1996;5:309–320.

20. Cella D, Hernandez L, Bonomi AE, et al. Spanish language translation and initial validation of the functional assessment of cancer therapy quality-of-life instrument. *Med Care* 1998;36:1407–1418.
21. Lent L, Hahn E, Eremenco S, Webster K, Cella D. Using cross-cultural input to adapt the Functional Assessment of Chronic Illness Therapy (FACIT) scales. *Acta Oncol* 1999; 38:695–702.
22. Eremenco SL, Cella D, Arnold BJ. A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Eval Health Prof* 2005; 28:212–232.
23. Altman BM, Gulley SP. Convergence and divergence: differences in disability prevalence estimates in the United States and Canada based on four health survey instruments. *Soc Sci Med* 2009;69:543–552.
24. Centers for Disease Control and Prevention. Prevalence and most common causes of disability among adults—United States, 2005. *MMWR Morb Mortal Wkly Rep* 2009;58:421–426.