Qual Life Res (2007) 16:175–186
DOI 10.1007/s11136-007-9190-8

Enhancing measurement in health outcomes research supported by Agencies within the US Department of Health and Human Services

Bryce B. Reeve · Laurie B. Burke · Yen-pin Chiang · Steven B. Clauser · Lisa J. Colpe · Jeffrey W. Elias · John Fleishman · Ann A. Hohmann · Wendy L. Johnson-Taylor · William Lawrence · Claudia S. Moy · Louis A. Quatrano · William T. Riley · Barbara A. Smothers · Ellen M. Werner

Received: 25 August 2006 / Accepted: 6 February 2007 / Published online: 26 May 2007

Abstract Many of the Institutes, Agencies and Centers that make up the US Department of Health and Human Services (DHHS) have recognized the need for better instrumentation in health outcomes research, and provide support, both internally and externally, for research utilizing advances in measurement theory and computer technology (informatics). In this paper, representatives from several DHHS agencies and institutes will discuss their need for better instruments within their discipline and describe current or future initiatives for exploring the benefits of these technologies. Together, the perspectives underscore the importance of developing valid, precise, and efficient measures to capture the full burden of disease and treatment on patients. Initiatives, like the Patient-Reported Outcomes Measurement Information System (PROMIS) to create health-related quality of life item banks, represent a trans-DHHS effort to develop a standard set of measures for informing decision making in clinical research, practice, and health policy.

This paper was written by authors in their capacity as employees of the US Department of Health and Human Services (DHHS), but the views expressed in this paper do not necessarily represent those of their Institute, Agency, or Center of DHHS.

B. B. Reeve (✉) · S. B. Clauser
Outcomes Research Branch, Applied Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, EPN 4005, 6130 Executive Blvd., MSC 7344, Bethesda, MD 20892-7344, USA
e-mail: reeveb@mail.nih.gov

L. B. Burke
Food and Drug Administration, Rockville, MD, USA

Y.-p. Chiang · J. Fleishman · W. Lawrence
Agency for Healthcare Research and Quality, Rockville, MD, USA

L. J. Colpe
National Institutes of Health (Currently at Substance Abuse and Mental Health Services Administration), Rockville, MD, USA

J. W. Elias
National Institute on Aging, Bethesda, MD, USA

A. A. Hohmann · W. T. Riley
National Institute of Mental Health, Bethesda, MD, USA

W. L. Johnson-Taylor
Division of Nutrition Research Coordination, Bethesda, MD, USA

C. S. Moy
National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA

L. A. Quatrano
National Center for Medical Rehabilitation Research, Bethesda, MD, USA

L. A. Quatrano
National Institute of Child Health and Human Development, Bethesda, MD, USA

B. A. Smothers
National Institute of Nursing Research, Bethesda, MD, USA

E. M. Werner
National Heart, Lung, and Blood Institute, Bethesda, MD, USA

Keywords Item response theory · Computerized-adaptive testing · Patient-reported outcomes · Health-related quality of life

Introduction

Many of the Agencies and Institutes that make up the US Department of Health and Human Services (DHHS) have recognized the need for better instrumentation in health outcomes research, and provide support, both internally and externally, for research to enhance measures of health-related quality of life (HRQOL) utilizing advances in measurement theory and computer technology (i.e., informatics). In this paper, representatives from several US DHHS Agencies and Institutes will discuss their need for better instruments within their discipline and describe current or future initiatives for exploring the benefits of these technologies for improved decision making in research, practice, and population surveillance.

National Institutes of Health (NIH)

Under the NIH Roadmap Initiatives aimed at fostering collaborations across the NIH Institutes and Centers in order to improve the clinical research enterprise [1], a trans-NIH effort is taking place to develop a public domain web-based resource, the Patient-Reported Outcomes Measurement Information System (PROMIS), that will measure key health symptoms and HRQOL domains that are relevant to a variety of chronic diseases [2]. Started in August 2004, the PROMIS initiative establishes a collaborative partnership between the NIH and multiple research sites, through a cooperative agreement (U01) mechanism, to develop the measurement system. The aims of PROMIS are to: (1) electronically administer both short form instruments and individually tailored patient-reported outcome (PRO) questionnaires (i.e., computerized adaptive testing, CAT) via a number of secure delivery platforms (e.g., computer, Internet, handheld, telephone); (2) collect PRO data for research and improvements to the system; and (3) provide instant health status reports to the patients, health care providers, and researchers.

Having a validated, dynamic system to measure patient-reported outcomes efficiently in study participants with a wide range of chronic diseases and demographic characteristics would greatly enhance the outcomes research enterprise and facilitate comparisons among research studies. Scientists will be better equipped to understand how patients perceive changes in their health status resulting from new treatments, thereby directing research to therapies that would be most highly valued by patients. Ultimately, this type of system will be useful in clinical practice to measure treatment response and guide therapy. Further, the ability to link electronic medical databases with systems like the PROMIS would strengthen our national capacity to monitor progress against the burden of disease and to support a wide range of studies on the determinants of health care utilization and outcomes [3].

One of the many PROMIS-supported research projects will identify and address challenges related to the implementation of PROMIS technology in multi-center clinical trials. This project will develop options for integrating computerized HRQOL measurement and data capture instruments, assess the readiness of sites in multi-center clinical trials to collect such data, conduct simulations to explore the costs and benefits for different data collection models, and develop recommendations for incorporating such systems in multi-center clinical trials.

National Cancer Institute (NCI), NIH

In 2007, an estimated 1,444,920 persons in the United States are expected to be diagnosed with cancer; about 559,650 are expected to die from cancer; and more than 10.5 million are undergoing curative treatment, coping with progressive disease, or living free of cancer after successful therapy [4]. The number of cancer survivors will continue to grow with recent advances in medicine and technology. Given these figures, addressing the effect of cancer symptoms on individuals' lives is becoming increasingly critical for efforts to reduce the burden of cancer and its treatment. Symptoms, such as pain, depression, and fatigue may persist or appear, even after treatment ends [5].

Substantial progress in reducing the suffering and death caused by cancer is being pursued by the NCI and cancer agencies and organizations worldwide through a variety of initiatives, programs, and projects. At the NCI, these efforts emphasize the joint importance of basic and applied scientific discovery, the development and testing of promising interventions, and the delivery of quality care to prevent, detect, and treat cancer and to improve the length and quality of life of cancer survivors [3].

The NCI is actively involved in exploring the state-of-the-science in cancer outcomes research and supporting the development of better instruments using modern measurement methods to enhance our ability to evaluate the impact of interventions and treatments on patients' symptoms and HRQOL. In 2002, an NIH state-of-the-science conference identified a need for increased assessment and monitoring of pain, depression, and fatigue in cancer patients, and a lack of quality instruments applicable for children and adolescents, older adults, individuals with

cognitive impairments, and individuals from different ethnic and cultural groups [5]. Panelists encouraged taking advantage of the advances in measurement theory and information technologies to develop and administer such instruments.

The NCI established in 2001 the Cancer Outcomes Measurement Working Group (COMWG) [6]. Comprising 35 experts drawn from academia, government, industry, and the cancer patient and survivorship communities, the COMWG was charged with evaluating the state of the science in outcomes measurement and recommending approaches to improve the scientific quality and usefulness of measures. Among many issues, the COMWG identified a need for outcome measures that are valid, reliable, and feasible across the continuum of cancer care from diagnosis and treatment to survivorship and end-of-life care [7]. The COMWG recognized the strengths of modern measurement theory for head-to-head comparison studies of leading HRQOL instruments to evaluate their strengths and weaknesses and the development of cross-walk tables to compare or combine results from studies using different instruments. Further, they recognized the potential value for item banking to develop both short form instruments and CAT that can be used in a variety of applications to enable HRQOL and other patient outcomes to be used more readily for medical and policy decisions.

To enhance implementation of item response theory (IRT) modeling in health outcomes research, the NCI funded two Small Business Innovations Research (SBIR) contracts, in 2004, to develop user-friendly IRT software tailored for health outcomes researchers that also increases flexibility, sophistication, and capabilities for users. The software will integrate multiple IRT applications such as item evaluation and scoring, DIF, and item banking in the same platform.

The NCI is actively engaged with the NIH PROMIS project [2]. In addition to the PROMIS network developing measures of generic HRQOL domains of physical function, pain, fatigue, emotional distress, and social/role participation, the NCI is supporting additional development of both short forms and CATs to measure sleep/wake function, perceived cognitive functioning, sexual functioning, and both positive and negative illness impact. This additional project enhances the PROMIS as well as improves the relevance of the system for cancer patients both in active treatment and survivorship phases.

One primary research area with expanded need for better HRQOL instruments are clinical trials. Over 1,500 NCI-supported cancer trials are conducted annually with many Phase III trials collecting patient self-reported symptoms and HRQOL information. The *NCI Strategic Plan for Leading the Nation to Eliminate the Suffering and Death Due to Cancer* [8] promotes the incorporation of HRQOL endpoints in NCI-supported clinical trials to assess the effectiveness of specific treatments and their influence on the quality of life for patients and survivors. This will involve the development of standardized measures of HRQOL and symptom severity for use in trials. Recognition of the need for psychometrically valid HRQOL instruments was highlighted in a 2006 NCI-sponsored conference, ''Patient-Reported Outcomes Assessment in Cancer Trials: Evaluating and Enhancing the Payoff to Decision Making. [9]'' More valid, precise, and sensitive instruments may reduce the sample sizes needed to determine a meaningful change or difference among treatment arms, and the use of item banks to develop a standardized set of instruments will facilitate the combining or comparison of study results.

Another key area that can benefit from improved instruments is in the care delivery setting. CAT-based instruments can be administered over a variety of devices such as telephone, Internet, or handhelds and this information on a patient's symptoms and HRQOL can be linked with a patient's medical record and monitored over time. The NCI is funding SBIR contracts to enable development of electronic HRQOL data collection systems for use by patients and oncologists in oncology practice and to demonstrate the benefits of such systems for patient–doctor communication and decision making. Such a system can instantly provide health status reports tailored for the patient or doctor and this information can be linked to guidelines and treatment recommendations.

Finally, NCI is incorporating HRQOL endpoints as part of its cancer control population-based surveillance systems. NCI has partnered with the Centers for Medicare and Medicaid Services to link SEER registry data with the Medicare Health Outcomes Survey as an innovative way to monitor HRQOL outcomes of cancer patients and survivors enrolled in Medicare managed care plans [10]. This new data source will provide population-based outcomes data on tens of thousands of elderly cancer patients and survivors. Utilizing modern measurement methods, this new linked database can potentially support a range of ongoing outcomes research, quality improvement, and health policy investigations related to cancer care delivery and HRQOL.

Division of Nutrition Research Coordination (DNRC), NIH

Two-thirds of American adults are either overweight or obese and almost one third of children are either at risk for overweight or obese [11]. Recognition of this growing epidemic and its potential impact on our youth has resulted in the development of a myriad of interventions aimed at stemming the crisis. Despite the intense flurry of activity

we remain cautious about the potential success of these ventures since we know that success will only be evidenced by those interventions that are able to sustain long term weight loss which the obesity research community has yet to attain. This lack of long term sustainability of weight loss continues to be a source of great speculation by researchers.

Some researchers attribute the lack of success to the failure to translate study findings to real life situations, in other words the failure to translate efficacious study findings into effective interventions [12]. Other researchers speculate that the environment which is filled with lots of food and little opportunity for activity overwhelms any intervention thereby biasing the weight change outcomes toward the null [13]. Still others believe that there is a lack of public will to make the changes necessary to realize the success experienced from research [14]. More recently however, there is a growing sentiment that our failures result from our approaching obesity as a simplistic ''acute'' disease when in reality it is a very complex, multifactorial ''chronic'' disease that requires multi-level interventions in order to change its course [15]. The complexity and chronicity of obesity make the sole use of biomedical indicators as outcome indicators unrealistic.

As in the case of other chronic illnesses, the main goals of any interventions should not be only to produce longer life but to also have the quality of that longer life be as good as it can be. This approach is being endorsed by obesity researchers as they acknowledge the impact that obesity has on physical functioning, social functioning and vitality. As a result more researchers are incorporating HRQOL as a study outcome.

Numerous generic and obesity-specific instruments have been used to assess HRQOL changes that occur as a result of interventions. These instruments have a number of limitations which could be addressed by the methods highlighted in this special journal supplement. First results from studies using these instruments can only be generalized to the sample population who were the basis of the development. This would not be a problem if the population of obese individuals were homogenous. However since the obese population is a heterogeneous population that varies by gender, race, and income, this is particularly problematic. For example, research has shown that a mildly obese white female has a very different health experience than a mildly obese African American female [16]. Therefore it is reasonable to believe that HRQOL items developed with a white female sample will function differently when administered to African-American females of the same BMI. Differential item functioning (DIF) testing allows researchers to examine possible racial/ethnic bias within obesity HRQOL questionnaires. Findings will allow the instrument developer to either modify the item to minimize DIF if possible or eliminate the item if the DIF is very large. This is a particularly important fact when one considers that most of the HRQOL questionnaires have been developed and validated using the majority population resulting in questionnaires that may not be appropriate or relative to the minority population who are bearing the brunt of the obesity epidemic.

Another limitation is the inability to make comparisons across studies that use different HRQOL instruments. IRT methodology allows researchers the ability to link two or more questionnaires on the same metric. Ultimately, the promise of the PROMIS project will allow multiple HRQOL forms to be developed from its item banks to allow the comparison or combination of study findings. These advantages represent a small subset of potential benefits for use of modern measurement theory in weight loss outcome studies.

Despite the potential benefits of IRT it is recognized that IRT is not without limitations. It is possible however that IRT will supplement current theory to enhance our understanding of HRQOL. That said, the revival of IRT and the recent focus on HRQOL as an outcome measure for obesity studies offers a unique opportunity to further both areas of study.

National Institute of Child Health and Human Development (NICHD), NIH

Disability can result from congenital factors, physical injury, disease or just every day wear and tear. The number of disabled individuals in the US is increasing. It is estimated that approximately 54 million or one in five Americans has a disabling condition [17]. Many of these disabled individuals will undergo rehabilitation in the hope of improving health and/or restoring function. Many measurement instruments exist to determine the effectiveness of rehabilitation interventions, but there is no consensus as to which one is best [18].

Rehabilitation models vary as to the point at which outcome can be measured [19–21]. Not withstanding this is a need for improved analytic tools to determine how we measure patient rehabilitation outcomes. The importance of measurement was recognized and discussed at the 2003 Physical Disabilities through the Lifespan Conference by some of the foremost experts in the field. The National Center for Medical Rehabilitation Research (NCMRR), within the NICHD, encourages not-for-profits, universities, and businesses to implement state-of–the-art measurement approaches using technology to provide accessible information and reduce the respondent burden.

A stronger evidence base of rehabilitation interventions through improved measurement techniques is a high

priority goal for the NCMRR. Evidenced based rehabilitation research needs to have valid measures of relevant constructs to assess the efficacy of medical rehabilitation for individuals with disabilities. To this end, the NCMRR recently issued a request for applications seeking investigators capable of developing pilot measures that build upon and extend advances in such fields as test theory, computer science and telecommunications. As a result, a number of grants were awarded to develop improved measurement tools for examining participation and function and some examples are discussed below.

The first project focuses on developing a means to enhance measures of physical activity levels for individuals with disabilities. It is anticipated that the development of a dynamic physical activity measure, using CAT methodology, will assist clinicians in evaluating the long term outcomes of rehabilitation programs in increasing physical activity behavior and functional performance in physically disabled populations.

A second investigator is developing a new instrument which measures participation and environmental factors. This web-based instrument will be available to individuals who have mobility limitations or who use devices for moving in their environment. The results of this instrument will be used to examine the continuum of activities of people with mobility limitations.

Dynamic assessment of pediatric health and functioning is a third initiative. This investigator is attempting to develop a practical yet precise system to measure health status for children with chronic health conditions across a wide range of medical care settings and services. Such a web-based instrument for use in assessing pediatric health and function may eliminate the current fragmentation of generic instruments that are currently used across varying age and diagnostic groups and achieve a substantial improvement in measurement breath and practicality.

The PROMIS is a separate but related project that provides another opportunity to improve measurement of symptoms in adults and children. For medical rehabilitation clinicians, researchers and consumers it is hoped that PROMIS will contribute by (a) achieving consensus about the principal symptoms that rehabilitation affects; (b) developing better means of measuring symptoms in rehabilitation science; and, (c) increasing the availability of a cross walk between measures that can be used in rehabilitation clinical trials.

National Heart, Lung, and Blood Institute (NHLBI), NIH

NHLBI's research agenda spans the most prevalent to very rare diseases of the heart, lungs, and blood [22]. In 2003, 23.5 million non-institutionalized adults (11%) in the US had a diagnosis of heart disease [23]. The prevalence of asthma among non-institutionalized adults was 20.7 million (9.7%) [24] and among children was 8.9 million (23%) [25]. In contrast, the most prevalent blood disease, sickle cell anemia, affects approximately 72,000 people in the US [26]. Hemophilia affects 18,000 people, primarily males, in the US [27]. Other genetic blood disorders, such as thalassemia, Diamond-Blackfan anemia, and Fanconi anemia, are rarer, affecting less than 1,000 persons, and lack reliable prevalence figures for the US.

NHLBI-funded research in PROs reflects the prevalence of chronic diseases in its research agenda. The Coronary Artery Surgery Study enrolled 780 patients between 1975 and 1979, and was among the first studies to formally collect data on HRQOL [28]. Emotional status and HRQOL were assessed among patients with chronic obstructive pulmonary disease in a trial comparing oxygen inhalation therapy protocols [29]. An HRQOL instrument specific for sickle cell disease (SCD) was identified as a top research priority at the 2002 workshop, *Adults with SCD: Meeting Unmet Needs*, and has been long-awaited by consumer advocates and clinical investigators [30].

In September, 2005, the NHLBI launched the Sickle Cell Disease HRQOL Questionnaire Project (Schre-QOL). The need for a sickle cell-specific HRQOL instrument is a testament to NHLBI's success in basic and clinical research that has increased the life expectancy of SCD patients from the teens to over 40 years of age [26].

In the US, SCD is primarily a disease of African-Americans. The adult population is very diverse in terms of health status, socioeconomic status, geographic location and health care access. Stroke and silent infarcts are relatively prevalent among children, and affect future neurocognitive functioning. There is, therefore, a wide range of educational achievement and use of technologies such as computers, among adults with SCD. Although many patients succeed in social and work roles, pain, fatigue, and unpredictable sickle crises challenge patients on a daily basis

Schre-QOL is designed to measure the impact of sickle cell and its treatments on patients in domains that are important, possibly, unique to SCD, and with items that reflect patients' experiences. Focus groups, critical incident reports, and key informant interviews are the qualitative research methods being used to generate data for developing items; while analytic methods, based on IRT, will be used to assess the psychometric properties of items in draft versions of the questionnaire. Based preliminary data, interactions with the health care system and stigmatization are issues that may qualify as domains. For example, patients report emergency room visits for pain crises as major stressors: they are accustomed to being labeled as

drug-seekers, and being under-treated for pain and its vascular sequelae. Although these issues are typically addressed in satisfaction with care instruments, they so overwhelm patients' experiences, that they adversely affect HRQOL. It will be a challenge to incorporate health care issues and stigmatization into an SCD HRQOL questionnaire.

Schre-QOL is partnering with the NIH-sponsored PROMIS initiative. The PROMIS questionnaires will serve as a generic instrument to allow HRQOL comparisons of patients with SCD and other chronic health conditions. The CAT technology fielded by PROMIS should facilitate data collection in this population, but its acceptability and ease of application will need to be evaluated.

The Hispanic Community Health Study (HCHS) is a multi-site, interdisciplinary epidemiologic study in US Hispanic populations sponsored by the NHLBI and six other NIH Institutes, and launched in late 2006 [31]. HCHS goals include studying the prevalence and pathogenesis of disease, and the role of acculturation and risk factors as protective or harmful among Hispanics. Several interesting measurement issues are being addressed by the investigators. Existing dietary intake instruments were developed for Mexicans and will need to be modified for linguistic and cultural relevance for Cubans, Puerto Ricans, and Central American cohorts. Although questionnaires to assess acculturation exist, newly-developed ones need to be validated in the study populations. Within the Hispanic context, it is essential to integrate cultural factors into classic health behavior models. This study will develop instruments and examine how the Latino cultural concepts of *familialism* (the high significance placed on the family unit), *collectivism* (the importance of friends and extended family in helping to make decisions in health), *simpatia* (the need for smooth interpersonal relationships in which criticism and confrontation are discouraged), *personalismo* (the preference for relationships with members of the in-group), and *respeto* (the need to maintain one's personal integrity and allow for face-saving strategies) operate as additional components that may add predictive strength to any of the classical health behavior and motivation models established under mainstream American health psychology.

National Institute on Aging (NIA), NIH

NIA's mission is to improve the health and well-being of older Americans by support of high-quality research on aging processes and age-related diseases. Germaine to this mission, the assessment of HRQOL (and related measures of well-being) is expected to grow in importance in response to increased demands for information regarding health disparities (national, cross-cultural, ethnic/racial), evidence-based outcomes of treatment, and treatment cost by culture interactions. Whether as a core measure, mediator/moderator, or basic covariate, HRQOL and related measures are viewed as primary surrogates for the assessment of burden of disease and illness. The availability of short, yet reliable and valid questionnaires has enhanced the efficacy of HRQOL assessment; however we lack benchmarks to compare across international groups, ethnic groups, and disease and treatment groups

A NIA-sponsored project, with Dennis Fryback as PI, tackles this task by establishing the comparability of five HRQOL preference-based instruments [Short-form Health Survey—SF36v2; EuroQol EQ-5D; Quality of Well-Being scale (QWB), Health Utilities Index (HU12/3), Health and Activities Limitation index—HALex] and their sensitivity to treatment and illness duration [32]. This project generates normative data for common illnesses, gender, age groups, index sensitivity to recovery from illness, duration of illness and mode of survey. The overall goal is to establish successful cross-walks between the measures to allow the responses on one variable to predict the responses on another variable.

A comprehensive review of well-being relative to economics draws attention to the fact that organizations and nations can monitor well-being as an anchoring point for the expected outcomes of changes in social programs and economics [33]. Individuals high in well-being later earn higher incomes and perform better at work than people who report low well-being. Similar results have been observed for self-report of low neuroticism and lower mortality [34]. Well-being is related to health and longevity [33] but the pathways are far from being understood [35] and attempts to relate the protective effects of well being to biomarkers related to health have not provided straight-forward outcomes. A key to the link between well-being and biology may be in the resilience to adversity that the personal growth and engagement aspects of well-being provide (eudemonic well-being) relative to measures of well-being involving self-assessment of happiness and contentment (hedonic well-being) [35]. That is, the links between health and well-being may be mediated by life circumstances and the ability to respond to them

Another study found that assessments of well-being quite often rely on global evaluations of life satisfaction or happiness [36]. A survey method developed by Kahneman Krueger, Schkade, Schwarz, and Stone [36] is called the Day Reconstruction Method (DRM), which is an offshoot of experience sampling methods with a focus on examining time use and the affective quality of those experiences. Respondents revive memories of the previous day by constructing a diary of the sequence of episodes during that day. Specific questions are asked about each episode to

induce accuracy of the feelings experienced during that time. A goal of the DRM is to characterize the affective experiences associated with different activities. By example, episodes might be exercising before work, getting children off to school, commuting, or meeting with colleagues [36]. DRM completion times can range between 45 and 75 min.

Recent applications of the DRM method include comparison between well-being of women in Columbus, Ohio and Rennes, France where the structure of well-being was markedly similar and rank order correlation between 21 activities was $r = .94$. The French spent more time alone and rated time with others as a greater benefit. Affective disposition predicted 28% variance in experienced happiness for both groups and sleep quality was a strong predictor as well [36].

NIA, with cofunding from the National Center for Minority Health and Health Disparities (NCMHD), funds six Resource Centers for Minority Aging Research (RCMARs), each of which has a Measurement Core [37]. A primary mission of the RCMARS is to conduct research leading to greater standardization of measurement between diverse populations. There is clear evidence of differences in mortality, morbidity, and access to health care within and between diverse and minority populations when compared with each other and with majority America. However, it is unclear what proportion of these differences is attributable to ''real'' disparities and what proportion to differential interpretation of instruments, questions, and knowledge between groups. A special issue of *Medical Care* (Volume 44, Issue 11, Supplement 3, 2006) addresses these issues.

National Institute of Mental Health (NIMH), NIH

The mission of the NIMH is to reduce the burden of mental illness and behavioral disorders through research on mind, brain, and behavior. The public health burden from mental disorders is enormous. The World Health Organization's Global Burden of Disease study reported that mental disorders comprise four of the top five sources of premature death and disability in 15–44-years-old in the Western world, and unipolar depression is the fourth leading cause of death and disability worldwide [38]. Mental disorders such as depression, bipolar disorder, schizophrenia, and autism are serious and potentially life-threatening illnesses for which accurate screening procedures, diagnostic evaluations, and outcome measures are needed to develop and evaluate new treatments and prevention strategies.

Perhaps more than any other field of medicine, clinical research in mental health relies on patient reports of symptoms and associated impairments. Although promising research is being conducted into the underlying neuropathophysiology of mental disorders, there are currently no laboratory or physical findings that can serve as reliable markers for these syndromes. The development of structured clinical interviews such as the SCID [39] have greatly improved the reliability and precision of clinical diagnoses of mental disorders, but even these more standardized clinical interviews rely predominantly on patient reports to the clinician (i.e., patient report by proxy) to derive a diagnosis. As the field evaluates the clinical and research utility of dimensional versus categorical psychiatric diagnoses, reliable and valid patient reports will become increasingly important to psychiatric diagnosis [40–42].

Although there are numerous commonly accepted self-report scales available for measuring a range of mental health variables, some of these scales are antiquated and lack adequate documentation of item development, reliability, and/or validity based on current standards. Modern psychometric approaches such as IRT and CAT offer promising possibilities for improved measures of psychopathology and associated functioning that are highly reliable, efficient, and valid. The NIMH currently supports a number of projects using IRT and CAT methodology to refine existing scales or to develop new depression scales. The NIMH also is involved in the PROMIS network. One PROMIS primary research site, University of Pittsburgh, is collecting responses to the PROMIS HRQOL domains from outpatient psychiatric patients. The PROMIS offers the potential to develop measures that more accurately reflect the experiences of patients with mental disorders.

Reliable and valid measures of behavioral and emotional constructs are important not only to mental health research but also to the broader biomedical research community. Mental disorders, as well as various subclinical syndromes and psychosocial risk factors, are comorbid with and contribute to the development, course, and treatment of numerous other medical disorders including cardiovascular disease, diabetes, cancer, HIV, and neurological disorders [43]. Accurate measurement of mental disorders and related behavioral and emotional constructs has become increasingly important in a wide variety of biomedical research efforts.

The NIMH support for the development and evaluation of measurement tools extends beyond the assessment of symptomatology and includes support for research to improve the assessment of cognitive and functional impairments often associated with severe mental disorders. Recent examples of such support include the NIMH MATRICS project [44] and the NIMH Functional Assessment program [45]. Measurement research supported by the NIMH extends beyond patient report and includes other-reports, behavioral observations, performance measures, psychophysiological variables, biological

markers, technological innovations in measurement, and the integration of these various measurement modalities. The NIMH also encourages research on the impact of mental disorders on patient reports, including the effects of culture, stigma, cognitive impairments, and affective/motivational states on patient report measures. Advances in science are often preceded by advances in measurement. Therefore, novel and innovative mental disorders measurement research is critical to advancing the field of mental health and reducing the burden of mental disorders.

National Institute of Neurological Disorders and Stroke (NINDS), NIH

The burden of neurological disorders is enormous. NINDS estimates that 50 million Americans suffer from a neurological disorder. The costs to society in terms of medical care and lost productivity may exceed hundreds of billions of dollars each year; the impact on the HRQOL and well-being of individuals and families affected by these disorders is incalculable [46]. Among the over 600 disorders that afflict the nervous system, most occur very rarely, though some, such as stroke, epilepsy, Parkinson's disease, migraine, multiple sclerosis, ADHD, and autism are relatively common diseases. Few effective treatments for neurological disorders exist. NINDS currently targets more than one-third of its extramural research budget to clinical research projects. New therapeutic approaches, including pharmacologic agents, biologics, devices, surgery, physical therapy, psychiatric and behavioral interventions, and gene therapy are currently being evaluated in nearly 200 clinical trials of a wide variety of neurological disorders.

Due to the nature of most of the neurological disorders, many of which are chronic or progressive, most therapeutic approaches are targeted to reducing symptoms, limiting disability, and slowing disease progression. However, clinical trials in neurology commonly measure the efficacy of new treatments in terms of reducing mortality or other clinical events. Many of the traditional clinical or functional measures of disease status, such as tests of muscle strength or counts of seizure frequency, do not adequately represent the full scope of the impact of disease on an individual with a chronic neurological disorder. More subjective components of patients' functioning, such as social, psychological, and mental well-being, may be more important indicators of disease impact. Measurement of patient-centered outcomes is a particular concern in clinical trials, where small differences in clinical measurements or imaging results may not translate into important benefit to the patients.

The field of PROs and HRQOL assessment in neurology clinical trials is relatively young. Some aspects of HRQOL have been incorporated into many recent clinical trials in neurology, usually as secondary outcome measures. The literature on clinical trials and HRQOL reveals three basic approaches: HRQOL is sometimes characterized by assessment of symptomatology (e.g., pain, mood); it may be measured by standard generic instruments (e.g., the SF-36 or SF-12); and increasingly, new disease-specific instruments are being developed almost on a study-by-study basis. Because of the lack of consensus about the best tools or approaches to measuring HRQOL within or across studies or disease areas, it is not possible to compare the relative burden of various neurological conditions to each other or, more importantly, to compare the relative benefits of one treatment over another on the same patient-centered outcome. For example, in the field of stroke, at least 12 different HRQOL scales have been developed; current stroke clinical trials at NINDS incorporate any of a variety of generic and/or stroke-specific scales, with little overlap among studies. A similar situation exists in clinical trials for other neurological disorders, such as epilepsy, multiple sclerosis and migraine.

An additional issue is the apparent reluctance to design trials with the primary objective of comparing the effect of treatment on HRQOL, presumably because these outcomes appear to be too subjective, too hard to define concisely, too complex to administer, and too difficult to interpret. Additionally, there is a paucity of condition-targeted HRQOL surveys for persons with neurological diseases that are reliable, valid, responsive, and brief enough to be feasibly administered in the clinical trials setting.

NINDS is sponsoring a new initiative to develop a coordinated approach to defining and measuring HRQOL in neurological disorders. The objective of this initiative is two fold: to develop a core set of questions that will assess dimensions of HRQOL that are universal to patients with chronic neurological disease, and to identify additional concerns that may be specific to particular groups of patients defined by disease, age, or other factors. The resulting item banks, which will include items relevant to major categories of neurological disorders, will support development of short forms or CAT approaches that can be easily incorporated into the majority of definitive trials sponsored by NINDS. This initiative is coordinated with the PROMIS project to allow the examination of the burden of neurological diseases compared to other chronic diseases.

Additional initiatives focusing on PROs are also underway at NINDS, including research to develop new, concise measures of cognitive function for use in clinical trials and other clinical research. Coupled with valid, standard HRQOL measurement approaches, these measures will permit more comprehensive characterization of patient-centered outcomes in neurology trials.

National Institute of Nursing Research (NINR), NIH

The NINR is mandated to encourage and support research to understand and ease the symptoms of acute and chronic illness, to prevent or delay the onset of disease or disability or slow its progression, to find effective approaches to achieving and sustaining good health. In particular, NINR is interested in research on health promotion/disease prevention; HRQOL, including self-management, symptom management, and caregiving; health disparities; and end-of-life studies. These activities are at the heart of nursing practice, and research is essential to discover and establish the science necessary for the highest quality nursing practice.

Whereas medical research focuses on understanding the causes of disease and discovering new treatments and cures for these diseases, nursing research focuses on promotion of health, and prevention and management of disease. Using interdisciplinary and often biobehavioral interventions NINR research seeks to help individuals, families, and communities achieve the highest quality of life across the lifespan, including the end of life

Measuring symptoms and other self-reported outcomes is central for much of the research supported by the NINR. Having better standardized measures for these self-reported outcomes will greatly enhance our research. Examples of NINR research in which better instrumentation based on modern measurement theory would be useful include: HRQOL for caregivers, patients approaching the end of life, patients with a wide variety of chronic illnesses, patients facing natural declines associated with aging, cancer patients and survivors, post-transplant and post-trauma patients, stroke survivors, and patients living in poverty or in rural areas where access of care is restricted. Health outcomes measurement also would be useful in health promotion and prevention intervention studies, studies of patients at risk for disease states, investigations of adherence to treatment regimes, and evaluations of technology-based interventions,

The NINR has supported many centers and investigator-initiated projects with primary foci on issues of symptoms, other self-report health outcomes, and HRQOL. Some of the current exploratory and core centers funded by the NINR include centers on symptoms, symptom interactions and health outcomes; biobehavioral research; trajectories of aging and health care; symptom management in life-threatening illness; preventing and managing chronic illness in vulnerable populations; enhancing quality of life in chronic illness; health promotion and risk reduction in special populations; and self-management interventions for populations at risk.

The NINR is the lead institute at the NIH for research on end of life care and was the primary institute that sponsored the recent State of the Science Conference on Improving End of Life Care, along with the Office of Medical Applications of Research. The conference panel called for better measures of symptoms such as pain and fatigue and HRQOL [47].

Current program announcements sponsored by NINR include Research on Clinical Decision-Making in Life-Threatening Illness; Biobehavioral Methods to Improve Outcomes Research; Symptom Clusters in Cancer and Immune Disorders; Mechanisms, Models, Measurement, and Management in Pain Research; Chronic Illness Self-Management in Children and Adolescents; Parenting Capacities and Health Outcomes in Youths and Adolescents; and Improving Care for Dying Children and Their Families. Access to more psychometrically sound, standardized measures that will derive from the PROMIS initiative will greatly benefit proposals in response to these PAs.

Agency for Healthcare Research and Quality (AHRQ)

The mission of the AHRQ is to improve the quality, safety, efficiency, and effectiveness of health care for all Americans. One Agency focus is to improve health care outcomes through research, typically on the organization, delivery, and financing of health services. Given this focus, the Agency is interested in developing instruments that will facilitate assessment of the extent to which patients are receiving high-quality health care. In this regard, several major AHRQ initiatives include an outcome measurement component.

One approach to assessing the quality of health care is to obtain measures of patients' experiences of care, or their evaluation of and satisfaction with the care they have received. The Consumer Assessment of Healthcare Providers and Systems (CAHPS®) is a major AHRQ project that has developed several measures of patients' experiences with health plans and providers. The psychometric adequacy of these measures has been examined using classical test theory [48] as well as confirmatory factor analysis [49, 50]. In addition, the equivalence of measures across different patient groups (e.g., Medicare vs. privately insured) has been studied [51]. Reliability and validity studies have been conducted not only for the original CAHPS® survey, which focused on experiences with health plans, but also for newly developed surveys dealing with patients' experiences with inpatient hospital episodes and with provider groups or individual providers. Further, IRT methods were used to examine the performance of items in a CAHPS® survey designed to elicit parental perceptions of dental care delivered to children in publicly funded programs [52].

As part of its Evidence-Based Practice Centers (EPCs), the AHRQ supports systematic reviews of the efficacy of

clinical interventions for specific diseases. The EPCs review relevant scientific literature on clinical, behavioral, organizational and financing topics to produce evidence reports and technology assessments. One aspect of these reviews is an evaluation of the adequacy of commonly used measures of outcomes or quality of care. For example, a recent Evidence Report reviewed measures of the quality of breast cancer care in women [53]. The Report concluded that much more work was required to transform recommended indictors of quality care into reliable and valid measures of quality.

The AHRQ also supports several large-scale data collection efforts that provide information on utilization and costs of health care. These national resources enhance analyses of health outcome measures that are available in these data. The Medical Expenditure Panel Survey (MEPS) obtains data annually from a nationally representative sample of the civilian population. Psychometric analyses of components of MEPS data have been conducted. Fleishman and Lawrence [54] examined differential item functioning in the widely used SF-12 measure, which has been incorporated in MEPS since 2000. Under an AHRQ grant, Fisher is applying Rasch modeling to MEPS data on (1) children's behavioral problems and (2) quality of usual source of care. The MEPS also included, from 2000 through 2003, the widely used Euro-Qol (EQ-5D) measure of health preferences. In 2004 and 2005, the MEPS included measures of depressive symptoms (the PHQ2) and general psychological distress (the Kessler 6-item scale). The latter scale, which is also included in the National Health Interview Survey, includes questions asking how often a respondent experiences certain symptoms of psychological distress during the past 30 days; cut-points have been developed to distinguish persons experiencing serious psychological distress.

In summary, AHRQ supports a variety of research efforts focused on using advanced psychometric techniques to refine and develop outcome measures. The initiative in using IRT and related techniques in projects supported by AHRQ rests with researchers. The MEPS data are publicly available through the AHRQ website and are potential resources for researchers interested in using IRT and related techniques to advance the measurement of patient outcomes and quality of care [55]. The use of modern psychometric techniques to refine measures of quality of care clearly represents one of the potentially fruitful investments in addressing topics that are central to the missions of AHRQ.

Food and Drug Administration (FDA)

The mission of the US FDA is, in part, to protect the public health by assuring the safety, efficacy, and security of human and veterinary drugs, biological products, and medical devices. The FDA is also responsible for advancing the public health by helping to speed innovations that make medicines more effective, safer, and more affordable; and helping the public get the accurate, science-based information they need to use medicines to improve their health [56].

There is an urgent need to improve the efficiency and effectiveness of the clinical trial process, including trial design, endpoints, and analyses. More attention and innovation need to be applied to disease-specific trial design and endpoints intended to evaluate the effects of medical products. Evidence of drug effectiveness is deemed substantial for claims in product labels or advertising if supported by adequate and well-controlled studies using endpoints that reliably and validly measure the specific concept(s) claimed. Use of PROs in clinical trials has always had a key role in drug development because (1) some treatment effects like pain and fatigue are known only to the patient, (2) there is a desire to know the patient perspective about treatment effectiveness, or (3) the patient can provide a unique perspective beyond clinical based measures [57]. From 1997–2002, 30% of the new drugs approved by the FDA contained PROs in their labels. In the past decade, the FDA has approved six cancer drugs based, at least in part, on PRO instruments which showed that the drugs improved functioning or relieved symptoms, such as pain, difficulty swallowing, or dry mouth [58].

In February 2006, the FDA released draft guidance for Industry for use of PRO measures in medical product development to support labeling claims [57]. The purpose of this guidance is to explain how the FDA evaluates such PRO instruments for their usefulness in measuring and characterizing the benefit of medical product treatment as perceived by the patient. The adequacy of a PRO instrument as a measure to support medical product claims depends on its developmental history and demonstrated measurement properties. Methods from measurement theory play a key role in the development and analysis of such instruments. They help to examine the dimensionality of measured concepts, select items that yield information on a patient's health status, and test for cross-cultural biases in responses to PRO items. The draft guidance has received a lot of attention and public comment [59, 60].

The most critical consideration during FDA's review of the primary and secondary endpoints used by sponsors to measure treatment induced effects is the adequacy and availability of a clearly established concept of measurement, i.e., what is the instrument measuring and is it meaningful in some sense as a treatment benefit? Examples of other clinical trial design issues include attention to controlling for bias, quantification of meaningful effect sizes, calculation of study sample size to demonstrate

treatment effects of that size, and the relationships among all trial endpoints.

The FDA encourages demonstrations of the application and added-value of IRT-based instruments and CAT in the clinical trial setting to determine how to apply established measurement principles to endpoints. Discussions around these examples will provide the basis for informed FDA evaluations. It would seem appropriate to evaluate the performance of IRT strategies against other known PRO testing and measurement paradigms. For example, choose a few areas and demonstrate how it would work—only then can informative evaluations occur.

Summary

This paper reflects the perspectives of only a few of the many agencies that make up the DHHS; however, one can see the commonalities among the pieces to understand the growing role of psychometric theory and methods in health outcomes research. All authors noted the importance of including the patient's voice for informing decision-making in clinical research, practice, and policy. Psychometric methods, such as IRT, provide us the opportunity to develop valid and precise PRO instruments with reduced response burden. Advances in information technologies and the importance of creating interoperable data systems and standardized PRO measures open up our ability to efficiently collect data, integrate both the clinician-reported and patient-reported data, and act on the results. While support for developing disease-specific PRO instruments is occurring within DHHS agencies and institutes to increase their ability to capture the full burden of disease and treatment on patients, there also is a strong growing trend to work across agencies to create standardized measures to have a means of cross-talk and comparison of disease burden. The PROMIS project is the best evidence of this trans-Department collaboration. Public domain, widely-accessible PRO item banks will strengthen our national capacity to monitor progress against the burden of disease and to support a wide range of studies on the determinants of health care utilization and outcomes.

References

1. NIH Roadmap: Accelerating Medical Discovery to Improve Health [web site]. Retrieved November 27, 2006 from: http://www.nihroadmap.nih.gov..
2. Patient-Reported Outcomes Measurement Information System [web site]. Retrieved November 22, 2006 from: http://www.nihpromis.org
3. Lipscomb, J., Donaldson, M. S., & Hiatt, R. A. (2004). Cancer outcomes research and the arenas of application. *JNCI Monographs, 33*, 1–7.
4. American Cancer Society. (2007). *Cancer Facts & Figures 2007*. Atlanta (GA): American Cancer Society.
5. National Institutes of Health State-of-the-Science Panel (2004). National Institutes of Health State-of-the-Science Conference Statement: Symptom management in cancer: pain, depression, and fatigue, July 15–17, 2002. *Journal of the National Cancer Institute Monographs 32*, 9–16.
6. Lipscomb, J., Gotay, C. C., & Snyder, C. F. (Eds.) (2005). *Outcomes assessment in cancer: Measures, methods, and applications*. Cambridge, UK: Cambridge University Press.
7. Gotay, C. C., Lipscomb, J., & Snyder, C. F. (2005). Reflections on COMWG findings and moving to the next phase. In J. Lipscomb, C. C. Gotay, & C. F. Snyder (Eds.), *Outcomes assessment in cancer: Measures, methods, and applications* (pp. 568–583). Cambridge, UK: Cambridge University Press
8. National Cancer Institute. The NCI Strategic Plan for Leading the Nation to Eliminate the Suffering and Death Due to Cancer. NIH Publication No. 06–5773, January 2006, Bethesda, MD, Retrieved January 30, 2007 from: http://www.strategicplan.nci.nih.gov/
9. National Cancer Institute and American Cancer Society co-sponsored conference: Patient-Reported Outcomes Assessment in Cancer Trials: Evaluating and Enhancing the Payoff to Decision Making [web site]. Retrieved November 22, 2006, from: http://www.outcomes.cancer.gov/publications/workshops/proact/
10. Outcomes Research Branch, National Cancer Institute [website]. Retrieved November 27, 2006, from: http://www.outcomes.cancer.gov
11. Ogden, C., Carol, M., & McDowell, M., et al. (2006). Prevalence of overweight and obesity in the United States 1999–2004. *Journal of the American Medical Association, 295*, 1540–1555.
12. Bonow, R. O. (2003). The challenge of balancing scientific discovery and translation. *Circulation, 107*(3), 358–362.
13. Hill, J. O., & Peters, J. C. (1998). Environmental contributions to the obesity epidemic. *Science, 280*, 1371–1374.
14. Evans, W. D., Finkelstein, E. A., Kamerow, D. B., & Renaud, J. M. (2005). Public perceptions of childhood obesity. *American Journal of Preventive Medicine, 28*(1), 26–32.
15. U.S. Department of Health and Human Services. (2001). The Surgeon General's Call to Action to Prevent and Decrease Overweight and Obesity. [Rockville, MD]: U.S. Department of Health and Human Services, Public Health Service, Office of the Surgeon General. Available from: U.S. GPO, Washington.
16. Wee, C., Phillips, R., & Legedza, A., et al. (2005). Health care expenditures association with overweight and obesity among US adults: Importance of age and race. *American Journal of Public Health, 95*, 159–165.
17. U. S. Department of Health and Human Services, To Improve the Health and Wellness of Persons with Disabilities [website]. Washington, DC, Office on Disability. Retrieved January 18, 2006, from: http://www.hhs.gov/od
18. Fuhrer, M. J. (Ed.) (1997). *Assessing Medical Rehabilitation Practices The Promise of Outcomes Research*. Baltimore, MD: Paul H. Brookes Publishing Co.
19. Research Plan for the National Center for Medical Rehabilitation Research (1993). Bethesda, MD: NIH Publication No. 93–3509.
20. World Health Organization. (1980). International Classification of Impairments, Disabilities, and Handicaps. Geneva.
21. Brandt, E. N. Jr., & Pope, A. M. (Eds.) (1997). *Enabling America: Assessing the role of rehabilitation science and engineering*. Washington, DC: National Academy Press.
22. Peterson, C. M., & Harvath, L. (2004). Current and future commitment to hematology research. *Hematology/Oncology Clinics of North America, 18*, 1215–1234.
23. National Health Interview Survey. (2003). Summary Statistics for U.S. Adults, Appendix III, Table I.

24. National Health Interview Survey (2003). Summary Statistics for U.S. Adults, Appendix III, Table IV.
25. National Health Interview Survey (2004). Summary Health Statistics for U.S. Children, Appendix III, Table IV.
26. National Heart, Lung, and Blood Institute (2002). Sickle cell research for treatment and cure. Bethesda, MD: NIH Publication No. 02–5214.)
27. Centers for Disease Control and Prevention, U.S. DHHS [website]. Retrieved November 27, 2006 from: http://www.cdc.gov/ncbddd/hbd/hemophilia.htm/
28. Smith, H. C., Frye, R. L., & Piehler, J. M. (1983). Does coronary bypass surgery have a favorable influence on the qualify of life?. *Cardiovascular Clinics, 13*, 253–264.
29. Heaton, R. K., Grant, I., McSweeny, A. J., Adams, K. M. & Petty, T. L. (1983). Psychological effects of continuous and nocturnal oxygen therapy in hypoxemic chronic obstructive pulmonary disease. *Archives of Internal Medicine, 143*(1), 1941–1947.
30. National Heart, Lung, and Blood Institute. [website]. Workshop on Adults with Sickle Cell Diseases: Meeting Unmet Needs. Executive Summary and Recommendations. Retrieved November 27, 2006, from: http://www.nhlbi.nih.gov/meetings/scdmtg/execsum.htm/
31. Hispanic Community Health Study [website]. Retrieved February 2, 2007, from: http://www.cscc.unc.edu/hchs
32. The Health Measurement Research Group [website]. Retrieved February 2, 2007, from: http://www.healthmeasurement.org
33. Diener, E., & Seligman, M. E. P. (2004). Beyond money: Toward an economy of well-being. *Psychological Science in the Public Interest, 5*, 1–31.
34. Mroczek, D. K., & Spiro, A. (2003). Modeling intraindividual change in personality traits: findings from the normative aging study. *Journal of Gerontology B Psychological Science and Social Science, 58*(3), 153–165.
35. Ryff, C. D., Singer, B. H., & Love, G. (2004). Positive health: connecting well-being with biology. *Philosophical Transactions of the Royal Society of London. Series B. Biological Sciences, 359*, 1383–1394.
36. Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: the day reconstruction method. *Science, 306*, 1776–1780.
37. Resource Centers for Minority Aging Research [website]. Retrieved February 2, 2007, from: http://www.rcmar.ucla.edu/
38. Murray, C. J. L., & Lopez, A. D. (Eds.). (1996). *The global burden of disease and injury series: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020* (Vol. 1). Cambridge, MA: Harvard School of Public Health on behalf of the World Health Organization and the World Bank, Harvard University Press.)
39. Spitzer, R. L., Williams, J. B. W., Gibbon, M. & First, M. B. (1992). The Structured Clinical Interview for DSM-III-R (SCID): I. history, rationale, and description. *Archives of General Psychiatry, 49*, 624–629.
40. Aggen, S. H., Neale, M. C., & Kendler, K. S. (2005). DSM criteria for major depression: Evaluating symptom patterns using latent-trait item response models. *Psychological Medicine, 35*, 475–487.
41. Kraemer, H. C., Noda, A., & O'Hara, R. (2004). Categorical versus dimensional approaches to diagnosis: Methodological challenges. *Journal of Psychiatric Research, 38*, 17–25.
42. Kupfer, D. J. (2005). Dimensional models for research and diagnosis: A current dilemma. *Journal of Abnormal Psychology, 114*, 557–559.
43. Evans, D. L., Charney, D. S., & Lewis, L. et al. (2005). Mood disorders in the medically ill: Scientific review and recommendations. *Biological Psychiatry, 58*, 175–189.
44. Green, M. F., Nuechterlein, K. H., & Gold, J. M., et al. (2004). Approaching a consensus cognitive battery for clinical trials in schizophrenia: The NIMH-MATRICS Conference to select cognitive domains and test criteria. *Biological Psychiatry, 56*, 301–307.
45. National Institute of Mental Health. [website]. Program Announcement: Functional Assessment of People with Mental Disorders. Retrieved November 27, 2006 from: http://www.grants1.nih.gov/grants/guide/pa-files/PA-05-037.html/
46. National Institute of Neurological Disorders and Stroke. [website]. Retrieved November 27, 2006, from: http://www.ninds.nih.gov/
47. National Institutes of Health State-of-the-Science Conference Statement on Improving End-of-Life Care [website]. Retrieved November 27, 2006, from: http://www.consensus.nih.gov/2004/2004EndOfLifeCareSOS024html.htm/
48. Hays, R. D., Shaul, J. A., Williams, V.S., et al. (1999). Psychometric properties of the CAHPS 1.0 survey measures. *Medical Care, 37*(supplement), 22–31.
49. Marshall, G. N., Morales, L. S., & Elliott, M., et al. (2001). Confirmatory factor analysis of the Consumer Assessment of Health Plans Study (CAHPS) 1.0 core survey. *Psychological Assessment, 13*, 216–229.
50. Hargreaves, J. L., Hays, R. D., & Cleary, P. D. (2003). Psychometric properties of the Consumer Assessment of Health Plans Study (CAHPS) 2.0 adult core survey. *Health Services Research, 38*, 1509–1527.
51. Bender, R. H., & Garfinkel, S. A. (2001). Differences in the structure of CAHPS measures among the Medicare fee-for-service, Medicare managed care, and privately insured populations. *Health Services Research, 36*, 489–508.
52. Hays, R. D., Brown, J., Brown, L. U., Spritzer, K. L., & Crall, J. J. (2006). Classical test theory and item response theory analyses of multi-item scales assessing parents' perceptions of their children's dental care. *Medical Care, 44*(11 suppl. 3), S60–S68.
53. Moher, D., Schachter, H. M., Mamaladze, V., et al. (2004). Quality of breast cancer care in women. Summary, Evidence Report/Technology Assessment No. 105. Rockville, MD: Agency for Healthcare Research and Quality Publication No. 04-E030–1.
54. Fleishman, J. A., & Lawrence, W. F. (2003). Demographic variation in SF-12 scores: True differences or differential item functioning. *Medical Care, 41*, III-75–III-86.
55. Agency for Healthcare Research and Quality. Medical Expenditure Panel Survey (MEPS). [website]. Retrieved February 2, 2007, from: http://www.meps.ahrq.gov/
56. Food and Drug Administration, U. S. Department of Health and Human Services. Challenge and opportunity on the critical path to new medical products. Retrieved March 11, 2004, from: http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html/
57. Food and Drug Administration. U. S. Department of Health and Human Services. Guidance for industry. Patient-reported outcome measures: Use in medical product development to support labeling claims. Retrieved February 24, 2006, from: http://www.fda.gov/cder/guidance/5460dft.pdf/
58. Bren, L. (2006). The importance of patient-reported outcomes... It's all about the patients. *FDA Consumer, 40*, 27–32.
59. Dueck, A., Halyard, M. Y., Frost, M. H., & Sloan, J. A. (2006). Meeting on the FDA Draft Guidance on Patient-Reported Outcomes. Patient *Reported Outcomes Newsletter, 36*, 1–4.
60. Revicki, D. A. (2007). FDA draft guidance and health-outcomes research. *Lancet*, *369*, 540–542.