3. Uses of Evidence in Disability Outcomes and Effectiveness Research

ALAN M. JETTE and JULIE J. KEYSOR

Boston University

PROMOTING THE HEALTH AND WELL-BEING OF ALL Americans—including those with disabilities—has emerged as a national priority since the passage of the Americans with Disabilities Act (ADA) more than ten years ago. The ADA marked the first explicit national goal of achieving equal opportunity, independent living, and economic self-sufficiency for individuals with disabilities (Americans with Disabilities Act 1989). Its passage marked a growing recognition of the needs of people with disabilities. Achieving the ADA's goals, however, requires more than simply satisfying its specific provisions; it requires the careful management of the health needs of persons with disability (Patrick 1997).

People with disabilities represent an increasingly recognized target population whose health care needs can be addressed and, it is hoped, improved through health services research. Health services research aims to improve health and health care systems through research on the structure, processes, and effects of health services (Hadley 2000; Shortell 1998). Health services research examines the use, costs, quality, accessibility, delivery, organization, financing, and outcomes of health care services (Hadley 2000). In 1999, when Congress reauthorized the Agency for Healthcare Research and Quality (AHRQ), the chief federal agency supporting health services research, it specified that the AHRQ address health care for certain populations, including persons with disabilities.

The Milbank Quarterly, Vol. 80, No. 2, 2002
© 2002 Milbank Memorial Fund. Published by Blackwell Publishing,
350 Main Street, Malden, MA 02148, USA, and 108 Cowley Road,
Oxford OX4 1JF, UK.

One branch of health services research is outcomes and effectiveness research, which is defined as studies of the end result of medical care, or the effect of health care on the health and well-being of patients and populations (Epstein 1990; FHSR 1994). Outcomes and effectiveness research examines the impact of health care (both specific interventions like drugs, medical devices, and procedures and broader interventions like programs and services) on the health outcomes of patients and populations (AHCPR 1999; Mendelson, Goodman, Ahn, et al. 1998). The goal of outcome and effectiveness research is to provide scientific evidence to guide the choice of effective health care decisions made by all who participate in health care: patients, providers, policymakers, and third-party payers (Clancy and Eisenberg 1998).

Outcomes and effectiveness research is one of several categories of health services research that can gauge how well the health care needs of persons with disabilities are being met. According to a report from the Conference on the Science of Disability Outcomes Research sponsored by the Disability and Health Branch, National Center for Environmental Health of the Centers for Disease Control,

There is substantial evidence that people with disabilities use more health and health-related services than their able-bodied counterparts, and that timely access to such services has more direct and profound implications on their lives, their health, and their well-being. There also is substantial evidence, from various sources and for various reasons, that their service needs have not been met. (Andresen, Lollar, and Meyers 2000, S1–2)

Over the past decade the scientific literature on disability outcomes and effectiveness has grown substantially (Andresen, Lollar, and Meyers 2000; Brown, Renwick, and Nagler 1996; IOM 1991; Keith 1995). Several important questions have not been adequately addressed, however. For example, How much has outcomes and effectiveness research improved the health outcomes of persons with disabilities? How could outcomes and effectiveness research affect the health and well-being of persons with disabilities? What type of evidence is needed, and what are realistic expectations for disability outcomes and effectiveness research?

In this article we examine the potential of outcomes and effectiveness research for persons with disability—that is, the extent to which outcomes and effectiveness research meets the health and well-being

needs of persons with disability. To accomplish this, we address what disability outcomes research is, what its goals and expectations are, what kinds of evidence are needed to conduct disability outcomes and effectiveness research, and whether contemporary tools are adequate. Throughout our discussion, we suggest some research priorities for disability outcomes and effectiveness research.

Realistic Expectations for Disability Outcomes Research

Interest in evaluating the extent to which health outcomes and effectiveness research has and can serve the needs of persons with disabilities has led to the recent merging of disability and outcomes research traditions into a new field termed “disability outcomes research.” The intent of this emerging field is to combine the research on traditional disabilities and that on mainstream outcomes and effectiveness to describe the current state of science in disability outcomes and to suggest future directions and priorities (Andresen, Lollar, and Meyers 2000).

Disability outcomes research is viewed narrowly as outcomes of medical rehabilitation (Fuhrer 1997; Keith 1995) and broadly as outcomes influenced by factors outside rehabilitation, such as the role of environmental factors in participation in life activities (Andresen, Lollar, and Meyers 2000). The latter conceptualization resembles the operationalization of disability as the dynamic interaction between individuals and their environment (IOM 1991, 1997; Verbrugge and Jette 1994; NIDRR 2002). Although there are different perspectives on the scope of outcomes in disability outcomes research, combining disability research and outcomes and effectiveness research into one field of inquiry holds substantial promise for improving the health of persons with disabilities. In order, however, for this line of research to accomplish this goal expectations must be realistic.

In a review article on the impact and lessons of 115 outcomes and effectiveness research studies sponsored by the AHRQ between 1980 and 1997, Stryer and colleagues (2000) showed that by far the most frequently examined types of outcomes and effectiveness research were descriptive epidemiology and comparative effectiveness. Such studies have provided information about the natural history of diseases, the prevalence and/or incidence of disease, risk factors related to disease, and

comparative evaluations of treatment outcomes, diagnostic approaches, or other management approaches. The majority of outcomes and effectiveness research studies covered in this review, though, had little *direct* impact on health care policies, practices, and outcomes.

The Stryer analysis also illustrated the prominence of the biomedical orientation of health services research, as most of the research they reviewed was on describing, identifying, and treating specific diseases rather than addressing the health needs of broad target populations. This is not surprising given the funding structure of many U.S. federal funding agencies. But to improve the health of persons with disabilities, we need a better understanding of the organizational structures and processes of care that are related to better outcomes—outcomes that are meaningful to persons with disabilities.

The Stryer analysis is useful to keep in mind when contemplating the short- and long-term goals of outcomes and effectiveness research for persons with disabilities. What are realistic goals for disability outcomes and effectiveness research? Can researchers and policymakers avoid some of the shortcomings and criticisms of mainstream outcomes and effectiveness research conducted in the past decade on the needs and concerns of persons with disabilities? (Anderson 1994). To answer these questions, we start by considering the types of evidence needed for disability outcomes research.

Types of Evidence

Despite the growing interest in disability outcomes research, the types of outcome evidence that should be collected remain unfocused and at times confusing, perhaps because of the inconsistent and sometimes nonexistent definitions of terms used in such research. Not enough attention has been paid to conceptual and definitional issues regarding what outcomes are most relevant to disability outcomes research (Cella and Chang 2000; Erickson 2000; Hambleton 2000; Hays, Morales, and Reise 2000). Instead, terms and concepts are referred to throughout the literature without being carefully defined, leading researchers to “talk past one another” (Andresen and Meyers 2000; Brown and Gordon 1999).

From our perspective, the potential of disability outcomes research depends first on a thorough discussion of the relevant outcomes to pursue. This discussion should involve not only researchers from the disability

outcomes research community but also persons with disabilities, who have a unique and essential perspective on which outcomes are most relevant. After the most relevant health outcomes have been determined, the research community can then define and assess each outcome.

The literature contains several conceptual frameworks that can be used to explore potential outcomes to be included as evidence in disability outcomes research (Dijkers, Whiteneck, and El-Jaroudi 2000; IOM 1991, 1997; Patrick 1997; Simeonsson, Lollar, Hollowell, et al. 2000; Verbrugge and Jette 1994; Wilson and Cleary 1995). One is particularly promising: Patrick's Model of Health Promotion for People with Disabilities (1997), since it depicts four broad planes of outcome: Disabling Process, Opportunity, Total Environment, and Quality of Life (see figure 1).

Each plane of Patrick's model is clearly detailed. Disabling Process shows the theoretical progression from disease or injury to the restriction of activities. The classic disablement outcomes are represented in this plane: disease or injury, impairment, functional limitation, and activity restriction or disability. Opportunity shows those outcomes related to independent living, economic self-sufficiency, equality of rights or status, and full participation in community life. In Patrick's view, Opportunity "represents the interaction between the total environment of the individual at his or her particular stage of life course, and the disabling process" (Patrick 1997, 259). Total Environment includes the individual's biologic and genetic makeup, demographic characteristics (race, gender, age), lifestyle behaviors, health and social care systems, and physical and social characteristics of the environment. Patrick regards Quality of Life as a distinct outcome that includes people's perceptions of their position in life in the context of their particular culture and value system and in relation to their personal goals, expectations, standards, and concerns. Patrick suggests that quality of life for persons with disabilities is influenced by the other three planes.

Patrick's model is a useful guide for researchers trying to identify relevant outcome evidence for a particular disability outcome research study. Depending on the focus of the project, the researcher will need to address the impacts articulated in one or more of the model's four outcome planes: Disabling Process, Total Environment, Opportunity, and Quality of Life. Let us consider several examples to illustrate how the model could be used. Major stakeholders in the U.S. health care system, including persons with disabilities, would like to find a way of evaluating

FIG. 1. A model of health promotion for people with disabilities. From Patrick 1997, 258.

the impact of health care resources and organizational structures. From a policy perspective, for example, researchers might ask whether the facilities in rural areas are adequate to address the health needs of persons with disabilities. Are there disparities in relevant health outcomes among persons with disabilities living in rural versus urban areas, and if so, are the differences related to the distribution of services? The Opportunity plane of Patrick's model might be most appropriate for this line of research. Research on urban versus rural disparities or on regional differences in health outcomes might focus on independent living opportunities or full participation, as described in Patrick's model.

Another research question related to health services for persons with disabilities might be whether specific characteristics of delivery, such as treatment procedures, patient-provider communication, and patient education, lead to different outcomes. Here the outcome evidence might focus on differences in outcomes under the Disabling Process, such as disease severity, impairments, functional limitations, or activity restriction. Similarly, researchers developing, implementing, and evaluating rehabilitation interventions designed to prevent the progression of disability or to restore or mitigate a loss of function caused by chronic conditions might concentrate on outcomes within the Disabling Process. For instance, Ettinger and colleagues (1997), in a randomized controlled trial of 439 community-dwelling adults with knee osteoarthritis, examined the effects of exercise on impairments, function, and disability. Compared with a health education control group, those participants who did aerobic exercises had a higher oxygen uptake (i.e., peak VO_2), reported less pain, performed better on functional tasks like walking and climbing stairs, and reported less physical disability.

As another example, interventions might address "opportunity" by minimizing the disadvantages of not participating in community life owing to environmental barriers. This type of intervention may target the interaction of individuals within the context of the environment such as modifying an individual's work environment. For instance, changing the physical design of work stations, allowing more time or additional breaks, providing human or technical support, or improving the commute from home to work, including parking accessibility and public transportation, could decrease work disability among persons with disabilities (Allaire 2001). An outcome study in this context would focus on elements of the Total Environment.

Although outcomes such as disability, impairments, independent living, function, and quality of life are common in the health services and disability literatures, there is a lack of conceptual agreement about what precisely these or similar terms mean. Some authors use the terms *health status*, *quality of life*, and *function* interchangeably to refer to the same or similar health outcome (Barnett 1991; Carr, Thompson, and Kirwan 1996; Clancy and Eisenberg 1998; Lehman 1995). The meaning, however, of *health status* or *quality of life* might range from negatively valued aspects of life, such as death, to more positively valued aspects, such as social functioning or happiness. Clancy and Eisenberg, for example, define health-related quality of life as including health status, symptoms, and a patient's preferences, values, and functioning. To others, function, health status, and quality of life are distinct and independent concepts (Brown and Gordon 1999; Mor and Guadagnoli 1988; Patrick 1997; Smith, Avis, and Assmann 1999). To make matters even more confusing, some authors do not even supply definitions of the terms they actually use. In a review of 75 articles on the quality of life, Gill and Feinstein (1994), for example, reported that only 15 percent provided a conceptual definition of quality of life.

Recent research suggests that persons with disabilities may perceive health status and quality of life differently. In a metaanalysis of 12 chronic disease studies, Smith and colleagues (1999) examined how patients assess their quality of life and how they differentiate it from health status. They found that people with chronic conditions perceived quality of life and health status as distinct. When rating quality of life, patients emphasized their mental health over their physical functioning, whereas when rating health status, patients emphasized their physical health over their mental health.

We believe that clear theoretical conceptualizations and definitions of disability outcomes are crucial to this line of research. Patrick's model adequately addresses both of these factors and could be used to identify the types of evidence needed for disability outcomes research.

Another important research priority is to explore whether even comprehensive outcome models like Patrick's cover all the outcomes relevant to disability outcomes research. For example, although Patrick's model includes many outcomes that are relevant to the clinical effectiveness of health care policies, strategies, and interventions, it does not address the effectiveness of the interpersonal aspects of care. An evaluation of patients' expectations and for and satisfaction with their care might be

an important outcome in disability outcomes research. The effectiveness of interpersonal care has been widely discussed as important to quality of care, and reports of patients' experiences are included in the Health Plan Employer Data and Information Set (HEDIS), with a mandatory standard patient survey (CAHPS) in the most recent version of HEDIS (Campbell, Roland, and Buetow 2000; NCQA 1999). Qualitative grounded theory research methods may be useful in this line of research.

Do Existing Measures Provide Adequate Evidence for Disability Outcomes Research?

How well have the existing patient-level outcome instruments operationalized relevant outcomes for persons with disabilities? Does the current assessment methodology adequately gauge how well the needs of persons with disabilities are being met now and will in the future? Are new assessment tools needed for disability outcomes research? Should developing instruments be a high priority of AHRQ for outcomes research in this area?

A comprehensive review of disability outcome assessment is beyond the scope of this paper. Several reviews, however, were recently published as a supplement to the *Archives of Physical Medicine and Rehabilitation* (Andresen 2000; Andresen et al. 2000; Andresen and Meyers 2000; Cohen and Marino 2000; Dijkers et al. 2000; Gray and Hendershot 2000; Lollar, Simeonsson, and Nanda 2000; Meyers and Andresen 2000; Meyers, Andresen, and Hagglund 2000; Vahle, Andresen, and Hagglund 2000). These reviews are part of the proceedings of the Conference on the Science of Disability Outcomes Research sponsored by the Centers for Disease Control in 2000 and address outcome measurement issues on health-related quality of life, functional status, disablement, and outcomes for children and youth, depression, and social health.

Our discussion focuses on the SF-36, a patient-centered outcome instrument used worldwide (Ware 1993). We will highlight some of the limitations and challenges faced in using the existing tools to assess outcomes like those detailed in Patrick's model in regard to the outcomes research needs of persons with disabilities. In particular, we look at how well the SF-36 assesses each area of Patrick's outcomes model: Disabling Process, Opportunity, Environment, and Quality of Life.

The 36-item short form of the Medical Outcomes Study questionnaire (SF-36) was designed as a generic indicator of health status and health-related quality of life for use in population surveys and evaluative studies of health policy and as an outcome measure for clinical practice and research (McDowell and Newell 1996; Ware 1993; Ware and Sherbourne 1992). The SF-36 includes items organized into eight different scales designed to assess four different health outcome dimensions: behavioral functioning, perceived well-being, social and role disability, and personal evaluations (perceptions) of general health in general (Ware 1993). Table 1 summarizes the health-related phenomena captured by the eight SF-36 scales.

How well does the SF-36 assess key outcomes elements of Patrick's model? Let us look at each outcome domain to evaluate its utility as a comprehensive instrument for disability outcomes research. The Disabling Process outcomes, consisting of disease or injury states, impairments, limitations in function, and/or activity restriction, are fairly well covered in the SF-36. Twenty-one of the SF-36 items contain several elements of the disabling process. The ten-item SF-36 physical functioning scale, for example, includes items that assess the concept of activity restriction (e.g., vigorous activities, moderate activities, bathing or dressing oneself) and specific limitations in basic functions (lifting or carrying groceries, climbing several flights of stairs, climbing one flight of stairs, bending, kneeling, stooping, walking more than a mile, walking several blocks, and walking one block). Items in the two SF-36 role limitations scales assess the activity restriction dimension of the disabling process: "During the past 4 weeks, have you had any of the following problems in your work or other regular daily activities. . . . (A) Cut down on the amount of time you spent on work or other activities; (B) Accomplished less than you would like, (C) Didn't do work or other activities as carefully as usual." One of the pain items included in the SF-36 pain scale reflects symptoms of underlying impairments as defined by Patrick: "How much bodily pain have you had during the past 4 weeks?"; a second pain item refers to the restriction of activities: "During the past 4 weeks, how much did pain interfere with your normal work (including work both outside the home and housework)." Another two-item SF-36 social activity scale refers to the restriction of social activities: "During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups?" And "During the past 4 weeks, how much of the time has

TABLE 1
Summary of Health Phenomena Captured by SF-36 Scales as Applied to Patrick’s Model
of Health Promotion for People with Disabilities

Scale	Disabling Process				Quality of Life	Opportunity	Environment
	Disease/ Injury	Impairment	Function	Activity Restriction			
Physical Functioning			•	•			
Role, Physical				•			
Bodily Pain		•					
General Health							
Vitality					•		
Social Functioning				•			
Role, Emotional				•			
Mental Health					•		

your physical health or emotional problems interfered with your social activities (like visiting friends, relatives, etc.)?”

How well does the SF-36 represent the other outcome domains in Patrick's model? Do the SF-36 scales measure quality of life (which Patrick defines as people's perceptions of their position in life in the context of their particular culture and value system and in relation to their personal goals, expectations, standards, and concerns)? The SF-36 includes only a few dimensions of quality of life. The five general mental health items (e.g., Have you been a very nervous person? Have you felt so down in the dumps that nothing could cheer you up? Have you felt calm and peaceful? Have you felt downhearted and blue? and Have you been a happy person?) and the four vitality, energy, and fatigue items (e.g., Did you feel full of pep? Did you have a lot of energy? Did you feel worn out? and Did you feel tired?) come closest to Patrick's concept of subjective perceptions of a person's position in life and thus may be viewed as indicators of quality of life. This assessment appears consistent with Ware's analysis (1993) of correlations of SF-36 scales with a general measure of quality of life, as drawn from the General Psychological Well-Being measure (Dupuy 1984). This measure asks each respondent, "How happy, satisfied, or pleased have you been with your personal life?" Six response choices are offered, ranging from "Extremely happy, could not have been more satisfied or pleased" to "Very dissatisfied or unhappy most of the time." Ware reports that the correlations between this quality of life item and the SF-36 scales were positive and significant, ranging from a low of 0.19 for the physical function scale to a high of 0.60 for the mental health scale when tested in the general U.S. population (Ware 1993, 9:24–5). Ware interprets these findings as indicating that the SF-36 approach to assessing health status can be useful in interpreting scores and explaining their implications for both quality of life and general health (Ware 1993, 9:25).

In regard to outcomes research concerning persons with disabilities, the SF-36 approach to quality of life dimensions appears both narrow and general and therefore of limited utility. While the SF-36 includes patients' perceptions of the psychological domain of quality of life, it offers no information about perceptions of other, nonpsychological life domains, such as physical health, level of independence, social relationships, environment, and spirituality. But the outcomes of these nonpsychological domains may be critical to assessing the effectiveness of health care delivery to persons with disabilities. Although this narrow scope of

quality of life dimensions would be understandable in a generic health outcomes instrument, the SF-36's coverage of quality of life is limited—at least according to Patrick's operationalization of the term. Thus, the SF-36 may have questionable value as an indicator of quality of life for disability outcomes and effectiveness research.

In regard to Patrick's domains of Opportunity and Environment, there are no items on the SF-36 that address these domains. Therefore, the SF-36 would be an inappropriate outcome instrument for research on opportunity and environment.

Although in the past the SF-36 has been widely used in disability outcomes research, researchers have reported several problems with the SF-36 as a disability outcome instrument (Andresen, Gravitt, Aydelotte, et al. 1999). For example, the wording of the functional items has been found to be offensive to people who use wheelchairs, because several functional mobility items ask if he or she can "walk" a distance or climb a stairs; that is, it ignores wheelchair functional mobility. Several research groups have also reported problems with floor effects for the physical function scale (Brazier, Walters, Nicholl, et al. 1996; Kersten, Mullee, Smith, et al. 1999) and, in some applications, problems with ceiling effects for both the physical and psychological scales (Andresen, Rothenberg, Panzer, et al. 1998).

This admittedly limited analysis of only one outcome instrument, the SF-36, in relation to Patrick's model of outcome evidence for persons with disabilities nonetheless illustrates another challenge in advancing the disability outcomes research agenda. The range of relevant outcomes is very broad, as illustrated by the many dimensions of each area in Patrick's framework. No one outcome instrument, and certainly not a generic instrument like the SF-36, will likely be able to meet the needs for outcomes research on persons with disabilities. Patrick's model makes it explicit that the uses of evidence for persons with disabilities must include outcomes well beyond the traditional emphasis on the Disability Process, as seen in instruments like the SF-36.

Another research priority—which is beyond the scope of this article—is to evaluate whether existing assessment instruments are psychometrically adequate and feasible. We recommend that the mainstream outcome assessment methodology be systematically analyzed for its appropriateness to and feasibility for the relevant outcomes in disability outcomes research. If the current tools lack important outcome dimensions for persons with disabilities, then their development should be a high health

services research priority as well and an important foundation for future disability outcomes research.

Another, related research priority for disability outcome assessment is to move beyond the field's reliance on traditional fixed-form outcome instruments, which, however carefully chosen, present problems. One common problem with short-form instruments is the frequently encountered floor and ceiling effects, in which large numbers of individuals who complete these instruments score at either the top or the bottom of the range, thereby reducing the measurement's precision (Andresen, Fouts, Romeis, et al. 1999; Brunet, Hopman, Singer, et al. 1996; DiFabio, Choi, Soderberg, et al. 1997; Rubenstein, Voelker, Chrischilles, et al. 1998). In response to these concerns about inadequate measurement precision and inadequate coverage of important outcome domains, some researchers use more comprehensive outcome instruments, which lead to the frustration and fatigue of many subjects overwhelmed by large and burdensome batteries of instruments (Meyers 1999).

One promising solution to the measurement problems with traditional fixed-form instruments is to use both computer-adapted testing (CAT) (Ware, Bjorner, and Kosinski 2000) and item-response theory (IRT) (Hambleton 2000; McHorney and Cohen 2000; Weiss 1982). CAT and IRT techniques are currently being used in the development of a new generation of disability assessment instruments designed for rehabilitation outcomes research (Haley and Jette 2000).

CAT methodology uses a computer interface (or a computerized interview/clinician report) that is tailored to the patient's own ability. The basic purpose of the adapted test is to mimic what an experienced clinician would do. A clinician learns most when he or she directs questions at the patient's approximate level of proficiency; outcome items that are either too easy or too hard provide little information. An adaptive test first asks questions in the middle of the ability range and then only those questions based on the level of the patient's responses. In this way, fewer questions need to be asked (individual respondent, interview, or clinical judgment), and the ones that are asked provide more precise information regarding an individual's placement along a continuum of functional ability. CAT applications require a large number of items in any one functional area (item pools), items that consistently scale along a dimension of low to high functional proficiency, and rules guiding starting, stopping, and scoring procedures.

A significant challenge in the development of CAT models in health care applications is the need for large, representative data sets (Ware, Bjoner, and Kosinski 2000). CAT applications in disability outcomes research require large samples of persons responding to assessment items to establish item-and-response characteristic curves for the complex modeling used in CAT programs. If the test items are not well constructed, then any subsequent application will be problematic. In a CAT application, each item must be carefully considered, since a CAT application has fewer items and not every person is tested with the same ones.

The strategy of matching items to respondents has been used for decades to construct short and precise educational and psychological tests. CAT programs use a simple form of artificial intelligence that selects questions tailored to the test taker, shortens or lengthens the test to achieve the desired precision, scores everyone on a standard metric so that results can be compared, and displays the results instantly. Each test administration is adapted to the respondent's own abilities. For example, a person who is able to "walk a mile" is not asked to respond to a question about "walking 50 feet." In practice, this approach minimizes the number of items that are administered to an individual to obtain an estimate of functioning in any particular content area. However, these tests require computers and "modern" psychometric methods that have only rarely been applied to health questionnaires. Algorithms for CAT applications are built from extensive modeling (structure, ordering, and interrelationships among items) within a scale representing a functional concept.

The psychometric methods that make it possible to calibrate questionnaire items on a standard metric ("ruler") also yield the algorithms necessary to run the "engine" that powers CAT assessments. These statistical models estimate how likely the persons at each level of health are to choose each response to each survey question. This logic is reversed to estimate the probability of each health score from a particular pattern of item responses. The resulting probability makes it possible to estimate each person's score, along with his or her specific confidence interval. In principle, one can derive an unbiased estimate of an outcome, that is, an estimate without any systematic errors, from any subset of items that fits the model. The number of items administered can be increased to achieve the desired level of precision.

Most statistical models for estimating such item parameters can be traced to the work of Rasch (1980) or a second tradition, item-response

theory (IRT) (Hambleton and Swaminathan 1985). These models emphasize accommodating the data at hand and assume unidimensionality, that is, that the items included on a particular scale measure only one concept. IRT-based outcome assessments, along with CAT outcome assessments applied to disability outcomes research, take less time and are less burdensome to patients, more flexible across different research applications, more efficient and less costly to administer, and, where needed, more precise than conventional approaches. Thus, they may have considerable promise for advancing disability outcomes research.

Summary

Outcomes and effectiveness research holds considerable promise for better meeting the health care needs of persons with disabilities. While Patrick's Model of Health Promotion for People with Disabilities offers a useful initial framework for resolving some of the past conceptual confusion about health outcomes, more qualitative and theoretical work is needed to identify a comprehensive range of health outcomes relevant to the health care needs of persons with disabilities. Second, we recommend that a systematic review and critical analysis of existing health outcome instruments be conducted to identify current strengths and shortcomings for use in disability outcomes research. If the current tools are lacking or inadequate, their development and testing should become priorities for health services research. Third and last, researchers should move away from traditional fixed-form instruments and toward IRT-based outcome assessment and computer-adaptive testing methods, for assessment instruments that will be more flexible, feasible, and precise for use in disability outcomes research.

References

- Agency for Health Care Policy and Research (AHCPR). 1999. *The Outcome of Outcomes Research at AHCPR: Final Report*. Available at <http://www.ahcpr.gov/clinic/out2res/> (accessed January 9, 2002).
- Allaire, S.H. 2001. Update on Work Disability in Rheumatic Diseases. *Current Opinion in Rheumatology* 13:93–8.
- Americans with Disabilities Act of 1989. 1989. 104 Stat 327: 101–336, 42 USC 12101 s2 (a) (8).

- Anderson, C. 1994. Measuring What Works in Health Care. *Science* 263:1080–3.
- Andresen, E.M. 2000. Criteria for Assessing the Tools of Disability Outcomes Research. *Archives of Physical Medicine and Rehabilitation* 81(suppl. 2):S15–20.
- Andresen, E.M., B. Fouts, J. Romeis, and C. Brownson. 1999. Performance of Health-Related Quality of Life Instruments in a Spinal Cord Injured Population. *Archives of Physical Medicine and Rehabilitation* 80:877–84.
- Andresen, E.M., G.W.J. Gravitt, M.E. Aydelotte, and C. A. Podgorski. 1999. Limitations of the SF-36 in a Sample of Nursing Home Residents. *Age and Aging* 28:562–6.
- Andresen, E.M., D.J. Lollar, and A.R. Meyers. 2000. Disability Outcomes and Research: Why This Supplement, on This Topic, at This Time? *Archives of Physical Medicine and Rehabilitation* 81(suppl. 2): S1–4.
- Andresen, E.M., and A.R. Meyers. 2000. Health-Related Quality of Life Outcomes Measures. *Archives of Physical Medicine and Rehabilitation* 81(suppl. 2):S30–45.
- Andresen, E.M., B.M. Rothenberg, R. Panzer, P. Katz, and M.P. McDermott. 1998. Selecting a Generic Measure of Health-Related Quality of Life for Use among Older Adults: A Comparison of Candidate Instruments. *Evaluation and the Health Professions* 21: 244–64.
- Barnett, D. 1991. Assessment of Quality of Life. *American Journal of Cardiology* 67:41c–44c.
- Brazier, J.E., S.J. Walters, J.P. Nicholl, and B. Kohler. 1996. Using the SF-36 and Euroqol on an Elderly Population. *Quality of Life Research* 5:195–204.
- Brown, I., R. Renwick, and M. Nagler. 1996. The Centrality of Quality of Life in Health Promotion and Rehabilitation. In *Quality of Life in Health Promotion and Rehabilitation: Conceptual Approaches, Issues, and Applications*, eds. R. Renwick, I. Brown, and M. Nagler, 3–13. Thousand Oaks, Calif.: Sage Foundation.
- Brown, M., and W.A. Gordon. 1999. Quality of Life as a Construct in Health and Disability Research. *Mount Sinai Journal of Medicine* 66:160–9.
- Brunet, D.G., W.M. Hopman, M.A. Singer, C.M. Edgar, and T.A. MacKenzie. 1996. Measurement of Health-Related Quality of Life in Multiple Sclerosis Patients. *Canadian Journal of Neurological Sciences* 23:99–103.
- Campbell, S., M. Roland, and S. Buetow. 2000. Defining Quality of Care. *Social Science and Medicine* 51:1611–25.

- Carr, A., P. Thompson, and J. Kirwan. 1996. Quality of Life Measures. *British Journal of Rheumatology* 35:275–81.
- Cella, D., and C. Chang. 2000. A Discussion of Item Response Theory and Its Applications in Health Status Assessment. *Medical Care* 38(9, suppl. II):II66–72.
- Clancy, C., and J. Eisenberg. 1998. Outcome Research: Measuring the End Results of Health Care. *Science* 282:245–6.
- Cohen, M.E., and R.J. Marino. 2000. The Tools of Disability Outcomes Research Functional Status Measures. *Archives of Physical Medicine and Rehabilitation* 81(suppl. 2):S21–9.
- DiFabio, R.P., T. Choi, J. Soderberg, and C.R. Hansen. 1997. Health-Related Quality of Life for Patients with Progressive Multiple Sclerosis: Influence of Rehabilitation. *Physical Therapy* 77:1704–16.
- Dijkers, M.P.J.M., G. Whiteneck, and R. El-Jaroudi. 2000. Measures of Social Outcomes in Disability. *Archives of Physical Medicine and Rehabilitation* 81(suppl. 2):S63–80.
- Dupuy, H. 1984. The Psychological General Well-being (PGWB) Index. In *Assessment of Quality of Life in Clinical Trials of Cardiovascular Therapies*, eds. N.K. Wenger, M.E. Mattson, C.D. Furberg, and J. Elinson. New York: Le Jacq Publishing.
- Epstein, A. 1990. The Outcomes Movement—Will It Get Us Where We Want to Go? *New England Journal of Medicine* 323: 266–70.
- Erickson, P. 2000. Assessment of the Evaluative Properties of Health Status Instruments. *Medical Care* 38(9, suppl. II):II95–9.
- Ettinger, W.H., R. Burns, S.P. Messier, et al. 1997. A Randomized Trial Comparing Aerobic Exercise and Resistance Exercise with a Health Education Program in Older Adults with Knee Osteoarthritis: The Fitness Arthritis and Seniors Trial. *Journal of the American Medical Association* 277:25–31.
- Foundation for Health Services Research (FHSR). 1994. *Health Outcomes Research: A Primer*. Washington, D.C.
- Fuhrer, M.J. 1997. *Assessing Medical Rehabilitation Practices: The Promise of Outcomes Research*. Baltimore: Paul H. Brookes.
- Gill, T., and A. Feinstein. 1994. A Critical Appraisal of the Quality of Quality-of-Life Measurements. *Journal of the American Medical Association* 272:619–26.
- Gray, D.B., and G.E. Hendershot. 2000. The ICIDH-2: Developments for a New Era of Research. *Archives of Physical Medicine and Rehabilitation* 81(suppl. 2):S10–4.
- Hadley, J. 2000. Better Health Care Decisions: Fulfilling the Promise of Health Services Research. 1999 AHSR Presidential Address. *Health Services Research* 35:175–86.

- Haley, S., and A. Jette. 2000. Extending the Frontier of Rehabilitation Outcome Measurement and Research. *Journal of Rehabilitation Outcome Measurement* 4(4):31–41.
- Hambleton, R. 2000. Emergence of Item Response Modeling in Instrument Development and Data Analysis. *Medical Care* 38(9, suppl. II):II60–5.
- Hambleton, R., and H. Swaminathan. 1985. *Item Response Theory: Principles and Applications*. Boston: Kluwer Nijhoff.
- Hays, R., L. Morales, and S. Reise. 2000. Item Response Theory and Health Outcomes Measurement in the 21st Century. *Medical Care* 38(9, suppl. II):II28–42.
- Institute of Medicine (IOM). 1991. *Disability in America*, eds. A.M. Pope and A.R. Tarlov. Washington, D.C.: National Academy Press.
- Institute of Medicine (IOM). 1997. *Enabling America: Assessing the Role of Rehabilitation Science and Engineering*, eds. E.N. Brandt, Jr., and A.M. Pope. Washington, D.C.: National Academy Press.
- Keith, R.A. 1995. Conceptual Basis of Outcome Measures. *American Journal of Physical Medicine and Rehabilitation* 74:73–80.
- Kersten, P., M.A. Mullee, J.A. Smith, L. McLellan, and S. George. 1999. Generic Health Status Measures Are Unsuitable for Measuring Health Status in Severely Disabled People. *Clinical Rehabilitation* 13:219–28.
- Lehman, A.F. 1995. Measuring Quality of Life in a Reformed Health System. *Health Affairs* 14:90–101.
- Lollar, D.J., R.J. Simeonsson, and V. Nanda. 2000. Measures of Outcomes for Children and Youth. *Archives of Physical Medicine and Rehabilitation* 80(suppl. 2):S46–52.
- McDowell, I., and C. Newell. 1996. *Measuring Health: A Guide to Rating Scales and Questionnaires*. New York: Oxford University Press.
- McHorney, C., and A. Cohen. 2000. Equating Health Status Measures with Item Response Theory. *Medical Care* 38(9, suppl. II):II43–59.
- Mendelson, D.N., C.S. Goodman, R. Ahn, et al. 1998. Outcomes and Effectiveness Research in the Private Sector. *Health Affairs* 17(5):75–90.
- Meyers, A. 1999. *Enabling Our Instruments: Assuring Access to Health Care Research*. Proceedings of the National Conference on Health Statistics, August 3. Washington, D.C.
- Meyers, A., and E.M. Andresen. 2000. Enabling Our Instruments: Accommodation, Universal Design, and Access to Participation in Research. *Archives of Physical Medicine and Rehabilitation* 81(suppl. 2):S5–9.

- Meyers, A.R., E.M. Andresen, and K.J. Hagglund. 2000. A Model of Outcomes Research: Spinal Cord Injury. *Archives of Physical Medicine and Rehabilitation* 81(suppl. 2):S81–9.
- Mor, V., and E. Guadagnoli. 1988. Quality of Life Assessment: A Psychometric Tower of Babel. *Journal of Clinical Epidemiology* 41:1055–8.
- National Committee for Quality Assurance (NCQA). 1999. *HEDIS 1999 Reporting Set Measures by Domain*. Available at <http://www.ncqa.org/pages/communications/news/h99meas.htm> (accessed February 19, 2002).
- National Institute of Disability and Rehabilitation Research (NIDRR). 2002. *U.S. Department of Education IOSERS*. Available at <http://www.nidrr.org/new/announcements/nidrr-brochure.html> (accessed February 19, 2002).
- Patrick, D.L. 1997. Rethinking Prevention for People with Disabilities. Part I: A Conceptual Model for Promoting Health. *American Journal of Health Promotion* 11:257–60.
- Rasch, G. 1980. *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: University of Chicago Press.
- Rubenstein, L.M., M.D. Voelker, E.A. Chrischilles, D.C. Glenn, R.B. Wallace, and R.L. Rodnitzkey. 1998. The Usefulness of the Functional Status Questionnaire and Medical Outcomes Study Short Form in Parkinson's Disease Research. *Quality of Life Research* 7:279–90.
- Shortell, S.M. 1998. HSR: Annual Report to Our Readers and the Field, September 1, 1996 through August 31, 1997. *Health Services Research* 33:7–10.
- Simeonsson, R., D. Lollar, J. Hollowell, and M. Adams. 2000. Revision of the International Classification of Impairments, Disabilities, and Handicaps Developmental Issues. *Journal of Clinical Epidemiology* 53:113–24.
- Smith, K.W., N.E. Avis, and S.F. Assmann. 1999. Distinguishing between Quality of Life and Health Status in Quality of Life Research: A Meta-analysis. *Quality of Life Research* 8:447–59.
- Stryer, D., S. Tunis, H. Hubbard, and C. Clancy. 2000. The Outcomes of Outcomes and Effectiveness Research: Impacts and Lessons from the First Decade. *Health Services Research* 35:977–93.
- Vahle, V.J., E.M. Andresen, and K.J. Hagglund. 2000. Depression Measures in Outcome Research. *Archives of Physical Medicine and Rehabilitation* 81(suppl. 2):S53–62.
- Verbrugge, L.M., and A.M. Jette. 1994. The Disablement Process. *Social Science and Medicine* 38:1–14.
- Ware, J. 1993. *SF-36 Health Survey: Manual and Interpretation Guide*. Boston: Health Institute, New England Medical Center.

- Ware, J., J. Bjorner, and M. Kosinski. 2000. Practical Implications of Item Response Theory and Computerized Adaptive Testing. *Medical Care* 38(suppl. II):II73–82.
- Ware, J., and C. Sherbourne. 1992. The MOS 36-Item Short-Form Health Survey (SF-36), I: Conceptual Framework and Item Selection. *Medical Care* 30:473–83.
- Weiss, D. 1982. Improving Measurement Quality and Efficiency with Adaptive Testing. *Applied Psychological Testing* 6:473–92.
- Wilson, I., and P. Cleary. 1995. Linking Clinical Variables with Health-Related Quality of Life. *Journal of the American Medical Association* 273:59–65.

Acknowledgments: We would like to thank Mrs. Ginger Quinn for her assistance with this manuscript. This work was supported by grants H133P99004 and H113B990005 from the National Institute on Disability and Rehabilitation Research.

Address correspondence to: Alan M. Jette, Ph.D., Sargent College of Health and Rehabilitation Sciences, Boston University, 635 Commonwealth Avenue, Boston, MA 02118 (e-mail: ajette@bu.edu).