1994

The efficacy of functional assessments in rehabilitation medicine

Keith E. Swanson
University of North Dakota

Follow this and additional works at: https://commons.und.edu/pt-grad
Part of the Physical Therapy Commons

Recommended Citation
https://commons.und.edu/pt-grad/432

This Scholarly Project is brought to you for free and open access by the Department of Physical Therapy at UND Scholarly Commons. It has been accepted for inclusion in Physical Therapy Scholarly Projects by an authorized administrator of UND Scholarly Commons. For more information, please contact zeinebyousif@library.und.edu.
THE EFFICACY OF FUNCTIONAL ASSESSMENTS

IN REHABILITATION MEDICINE

By

Keith E. Swanson
Bachelor of Science in Physical Therapy
University of North Dakota, 1993

An Independent Study
Submitted to the Graduate Faculty of the
Department of Physical Therapy
School of Medicine
University of North Dakota

in partial fulfillment of the requirements
for the degree of

Master of Physical Therapy

Grand Forks, North Dakota
May
1994
This Independent Study, submitted by Keith E. Swanson in partial fulfillment of the requirements for the Degree of Master of Physical Therapy from the University of North Dakota, has been read by the Faculty Preceptor, Advisor, and Chairperson of Physical Therapy under whom the work has been done and is hereby approved.

(Faculty Preceptor)

(Graduate School Advisor)

(Chairperson, Physical Therapy)
PERMISSION

Title The Efficacy of Functional Assessments in Rehabilitation Medicine

Department Physical Therapy

Degree Master of Physical Therapy

In presenting this Independent Study Report in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the Department of Physical Therapy shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my Independent Study Report or, in her absence, by the Chairperson of the department. It is understood that any copying or publication or other use of this Independent Study Report or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my Independent Study Report.

Signature

Date

3/8/94
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................... v  
ABSTRACT ..................................................... vi  

CHAPTER  
   I. INTRODUCTION ........................................... 1  
   II. MEASUREMENT OF FUNCTION ....................... 5  
   III. PROBLEMS IN ASSESSING FUNCTION ............... 9  
   IV. A COMPARISON OF THE INDICES .................... 17  
      Introduction ........................................... 17  
      Barthel Index ......................................... 17  
      Katz Index of ADL .................................... 23  
      Functional Status Index .............................. 27  
      Summary of Physical Function Tests .............. 31  
   V. CONCLUSION ............................................. 34  

REFERENCES ................................................ 39
ACKNOWLEDGEMENTS

I wish to extend my appreciation to Mr. and Mrs. Brian Perreault.
ABSTRACT

Functional assessment in rehabilitation medicine is not a new concept. Functional assessment measures have, however, experienced increased usage. Health insurance companies and clinicians are now focusing on functional outcomes in addition to objective tests.

Certain measurement criteria must be satisfied before a measurement tool may be successfully implemented into the clinical setting. Variable standardization, reliability, and validity are necessary for accurate measurement. The term function, however, deals with many variables leading to definitional difficulty.

The purpose of the functional assessment is to describe, screen and assess, and monitor. Although many functional assessments are available for clinical use, most lack scientific rigor. Consequently, few reliability and validity studies have been completed. Therefore, a critical analysis of three common functional measurement scales was completed.

No scale demonstrated unequivocal superiority. Each scale has its own strengths and weaknesses. Scale selection depends on the clinical setting, the patient population, and the purpose for completing the assessment. Research in the area of definitional standardization, validity, and clinical feasibility is
necessary to further substantiate the efficacy of functional assessment in rehabilitation medicine.
CHAPTER I
INTRODUCTION

The concept of rehabilitation dates back approximately 50 years.¹ Rehabilitation has experienced great advances since such notables as Frank H. Krusen, Howard A. Rush, and other founders began to lay the framework of rehabilitation.¹ Throughout the years, rehabilitation has attained a wide array of closely related yet distinct definitions. Kottke and Lehmann¹ define rehabilitation as "a complex process of the integrated application of many procedures to achieve the restoration of the individual to his or her optimal functional status at home and in the community that the appropriate utilization of all of that patient's residual assets allows." Johnstone,² in 1978, defined rehabilitation as "obtaining the maximum degree of physical and psychological independence after disability by means of a carefully planned program."

Rehabilitation may also take on a specific meaning. Bobath,² for example, defines rehabilitation as "teaching the patient to manage his or her own life given the limitations of the damage to the CNS."

Although various professionals within the field of physical and rehabilitative medicine continue to selectively define rehabilitation, a common
denominator seems to remain consistent. Maximization of the individual's functional capacity is a central theme in the field of rehabilitation.

Like rehabilitation, the term function is also cited in the text as taking on different meanings depending on the content implied. A neutral definition of function from Dorland's Illustrated Medical Dictionary is: "the special, normal, or proper physiologic activity of an organ or part." Function in a specific sense can pertain to the performance of a body part, such as function of the shoulder. In contrast, function also describes the performance of an individual in a holistic sense, such as performance in activities of daily living (ADL). An operational definition of functional status is imperative to comprehend the broader, more general concept of health. Health, is defined by the World Health Organization as a state of complete physical, mental, and social well-being, not merely the absence of disease and infirmity. This definition may suffice for public usage; however, a more accurate and descriptive explanation is needed for the health care professional. Therefore, health is broken down into three parts: physical manifestations, symptoms of feeling states, and functional status. Thus functional status is represented as a major component part of the broader concept, health.

The physical therapist must acknowledge all aspects of health in order to set goals tailored to the patient. For example, a 16-year-old BK amputee does not merely want to learn how to walk with a prosthesis; he also wants to work on balance and coordination in order to play basketball. Susan O'Sullivan
states: "The patient and family should be considered active members of the team and should be closely involved in the goal setting process." "Without direct involvement, it will be difficult to motivate the patient and to focus attention on the attainment of these goals." She goes on to explain that many rehabilitation plans have failed terribly because the patient had established a very different set of goals than the health professional.

Jette breaks functional status into four subparts: physical function, emotional function, mental function, and social function. The first subpart is physical function which includes walking, climbing stairs, cooking, and generally getting around in the home environment. Physical therapists are trained to assess and treat these concerns. Second, functional status includes emotional function which pertains to an individual's ability to cope effectively with life stressors. The third subpart is mental function and includes intellectual, cognitive, and reasoning capabilities of the individual. Finally, social function includes interactions with other people and performance in social roles or obligations.

All professionals working in the rehabilitation field have differing educational backgrounds. Consequently, all must work together to achieve the common goal of maximal functional status as described above. The Occupational Therapist may focus primarily on physical problems of the upper extremity but may also recognize mental deficits which will affect the patient's physical performance of both basic ADL and the more complex instrumental
ADL. A Psychologist is available to assess mental capabilities of the patient. The Social Worker focuses primarily on social ability and may also assess affective function. As mentioned previously, the Physical Therapist is primarily concerned with the physical performance of skills needed to attain a maximal level of independence. However, just as other members of the rehabilitation team are concerned with different dimensions of the patient's functional status, so too is the physical therapist. Therefore, it is logical that effective communication across disciplines is imperative. Effective, meaningful communication between health professionals demands a need for measurement of functional status within the rehabilitation setting.

There are a myriad of functional assessment measurement tools in use today. Global assessments tend to focus on the larger concept of health status as defined earlier in the text. Examples of commonly used global or health status measures include the Arthritis Impact Measurement Scale (AIMS), the Sickness Impact Profile (SIP), and the Bush Index of Well Being. In contrast to global assessments, functional status measures are more uni-dimensional because they are only concerned with a patient's physical function level. Commonly used functional assessments include the Barthel Index,9,10,11 Katz Index of Daily Living,12,13,14 and the Functional Status Index.15,16 Functional status measurement instruments attempt to measure items clinically relevant to the physical therapist. Thus, the main purpose of this paper is to critically examine functional status measures and define their role in health care delivery.
CHAPTER II
MEASUREMENT OF FUNCTION

An extensive amount of research has been completed to determine the efficacy of functional status assessments. All assessments deal with attempting to describe function quantitatively, in measurable terms. A closer look at the term measurement as it pertains to function is necessary.

Michels\textsuperscript{17} states that measurement in physical therapy has long been neglected. Successful research depends on sound knowledge of formal measurement theory. Measurement can therefore be viewed as a means to the end of achieving the science of Physical therapy. Jette\textsuperscript{7} also points out that without a scientific basis for the assessment and measurement process, the physical therapist's ability to function as an independent practitioner in terms of communicating with fellow professionals, documenting treatment efficacy, and claiming scientific credibility for the profession is impossible. Many different definitions of measurement are offered in the literature. Michels\textsuperscript{17} states: measurement is the act of converting, counting, ranking, and quantifying. Stevens,\textsuperscript{18} recognized as one of the originators of modern measurement theory, states that in the broadest sense, measurement is the assignment of numerals to objects or events according to rules.
The variable or dimension to be measured must be identified before a particular measurement can take place.\textsuperscript{19} This is known as an operational definition.\textsuperscript{19,7} The operational definition is quite self explanatory when discussing dimensions of strength, joint ROM, or tone. These variables are universal and therefore can be understood and used by all similarly trained people. These particular variables also possess sound theoretical assumptions.\textsuperscript{7} A variable which is universal and possesses sound theoretical assumptions is publicly observable.\textsuperscript{19} A variable without an operational definition (publicly observable) cannot be measured.

After an operational definition has been delegated to a particular dimension, the next step involves operationally defining two or more units or categories of that dimension.\textsuperscript{19} Depending on the particular dimension being measured, categories under that dimension may range from two to infinity. For example, the most simple functional assessment index may classify patients as being either independent or dependent. Notice that with this theoretical example, the two categories are exhaustive; the patient is either independent or dependent. All patients are included within these operationally defined parameters. Of course, not all and, in fact, very few measurements have only two categories.\textsuperscript{19} An example is the manual muscle test which has six categories: zero, trace, poor, fair, good, and normal. Regardless of the number of categories, they must be mutually exclusive. This will avoid the ambiguity of not knowing which category a patient should be assigned.\textsuperscript{19} Due
to the multifaceted and complex dimension of functional status, many problems have been cited when attempting to measure functional status. 

There are many different functional status measures in use today. The operational definition of function is not standardized. This can lead to confusion across rehabilitative disciplines and insurance companies as to what is truly being measured. An attempt to standardize definitions in health care was undertaken by the World Health Organization in 1980. The International Classification of Impairments, Disabilities, and Handicaps (ICIDH) precisely classifies the consequences of disease. The ICIDH defines impairment as abnormalities of body structure and appearance. In essence, impairments describe abnormalities at the organ level. Disability is any restriction or lack of ability (resulting from an impairment) to perform an activity in the manner considered normal for a human being. Whereas impairments are concerned with individual functions of the parts of the body, disability is concerned with compound or integrated activities expected of the person such as various tasks, skills, and behaviors. Wood and Badley state that disability is used to describe a functional departure from the norm. When the term disability is used within the context of health care, it is considered the loss or reduction of functional ability and activity that is dependent upon a particular impairment. A handicap takes into account the value attached to an individual’s status or performance when it departs from a norm. Therefore, handicaps depend on cultural norms. Two individuals with the exact same disabilities may be considered handicapped by one group of people but not by the next.
So where does functional status or functional limitations fit into the consequence of a particular disease? The answer to this question is still not crystal clear. The ICIDH\textsuperscript{20} states that functional limitations fit under the category of impairments. Yet Granger and Koshel\textsuperscript{22} state: "In the way that analysis of pathology defines the nature of impairment, analysis of functional limitations defines the nature of disability." Although the above definitions serve as a foundation for standardizing the meaning of disease and its consequences, complete agreement has not been achieved. The definitions will apply throughout the remainder of this paper. The author hopes these standardized definitions will minimize the inherent subjectivity of terms like functional disability.
CHAPTER III

PROBLEMS IN ASSESSING FUNCTION

The assessment of "function," using standardized functional measurement tests, has experienced many difficulties over the years. Wood and Badley published a study stating that approximately 34% of Great Britain’s population is impaired in some way. The extent to which these impairments affect the individual’s functional status varies. A particular impairment may result in severe loss of function for one individual, while the same impairment may lead to little if any loss of function for someone else. For example, a right humeral fracture will mean much greater loss of function for a baseball pitcher verses a business manager. Functional status measures attempt to define the scope of benefit provided by medical intervention. A number of factors can affect the potential benefit of the assessment. These factors include, but are not limited to, personal circumstances of the individual, severity of the underlying disorder, services available at the time, and society's commitment to meeting certain functional levels. These variables will affect how well a functional assessment index approximates the actual functional limitation.

The purposes of a functional assessment measure are numerous. Granger indicates the functional assessment and subsequent analysis of the information provided by a functional assessment are useful for: planning treatment, determining
effectiveness and efficiency of treatment, maintaining continuity of care, developing treatment resources, and also improving treatment resources. Jette simplifies the purposes of the functional assessment into three areas. They are to describe, to screen and assess, and to monitor. Description involves the collection of information or data to establish a body of information about the concept under investigation. Functional assessments geared toward describing are used primarily in the research community to compile normative data for study comparisons in research. Screening and assessing pertains to a detailed review of physical function leading to conclusions about the nature of the problem. Functional assessments which aim to screen and assess must be more detailed than a descriptive measure and, as a consequence, are frequently multidimensional and more time consuming to administer. Monitoring refers to detecting a change in function through repeated measurements over time. A functional assessment designed to monitor must be sufficiently precise to detect the level or degree of change anticipated.

Many functional status assessments are available to the practicing clinician. The Physical Therapist must determine the necessity for and the type of measurements needed. Unfortunately, a universal, all encompassing assessment measure to document functional outcomes is not available. The feasibility of this goal is questioned by some researchers. Liang, Cullem, and Larson state that the process of measuring functional outcomes is probably not possible. A number of reasons why perfection is thought to be unattainable follows.
The first reason is that no instrument is perfect in terms of basic psychometric criteria. Liang and Jette reviewed 12 common functional status measures. They describe five criteria which must be fulfilled to provide an adequate, accurate measurement of function. The criteria were: 1. the ability to permit quantification, 2. sound validity, 3. sound reliability, 4. standardization of data collection, and 5. ability to distinguish changes which are adequate for its intended purpose. Twelve functional assessments were tested along the five criteria. Surprisingly, not one fulfilled all five criteria. Moreover, not even 50% satisfied basic criteria of reliability and validity in scale construction.

A second reason for unattainable perfection in a functional status assessment involves the choice of specific items in the makeup of the health status scale. Because not all costs or benefits are measurable, the selection of outcome items from the infinite possibilities must be carefully thought out. For example, how could any assessment measure the effect of prednisone when long term side effects, such as chondromalacia, has not yet materialized into function limiting symptoms?

The inability for functional capacity to be absolute is the third obstruction for attaining a perfect assessment. Functional capacity is relative to the patient’s expectations, priorities, goals, social supports, and other factors. Implicit utility weighting has been used to give certain items more importance in an effort to judge the desirability of each component or level of function. The more important an item is thought to be, the more weight that item will be assigned. Implicit utility weighting is a difficult process. Disagreement exists in variable priority as well as individual
differences. How does one weight the relative importance to an individual versus
groups of individuals versus society? 

A fourth reason that Liang, Cullem, and Larson believe it is impossible to
construct a perfect functional assessment is because the measurable drives out the
important. Alternatively stated, blind faith in the mathematics of scale construction
may lead clinicians to down-play the importance of human clinical judgment of
functional status.

The moral implications involved in attempting to measure health status also
posses difficulty in scale construction. Faden and Leplege believe that quality of life
is a morally loaded term. As researchers attempt to quantify such a term, they also
need to keep in mind certain ethical issues. Perhaps the most basic question is: What
gives quality to life? Every person has their own perception of a quality life. Unique
interests and enjoyments make every individual’s answer to this question unique.
Consequently, certain human states are considered universally undesirable. Pain,
physical incapacitation, or depression are all undesirable. Similarly, other states are,
for the most part, considered desirable. Physical mobility, peace of mind, and feeling
loved are all states that individuals value and are desirable in the search for a ’good’
quality of life. An important point that needs to be recognized is that even with
respect to these states, individuals vary as to the relative value or importance they
attach to each. The quality of life for one individual may be viewed through one’s
eyes as being poor, while that same quality of life may look good when seen through
another’s eyes. Quality of life really depends on the individual’s personality, desires,
and history. Faden and Leplege\textsuperscript{5} suggest that caution should be exercised when quality of life or health status assessments are used because a standardized questionnaire is probably not the patient’s idea of the moral commitments of medicine. Patient’s believe and expect validation of their fears and the offer of hope when medical help is sought. No high technological assessment can ever take the place of the time honored question of: "How are you feeling today?"

There are a number of shortcomings and obstacles preventing the existence of a perfect functional assessment scale. The use of functional scales in the clinic is quite infrequent because of a number of factors. Golden\textsuperscript{25} has attempted to explain reasons for implementation difficulty. The incorporation of any new clinical intervention into routine clinical practice requires three fundamental steps: exposure, learning, and adoption. Golden further explains that functional health measurements have not yet passed through any of these steps. Golden believes the only way functional health measures will gain acceptance in clinical practice is through persuasion of key pacesetter physicians. This may be accomplished by the introduction of health status measurements directly into current medical school curricula as well as paraprofessional curricula where young medical professionals’ attitudes toward new innovations such as functional assessments are shaped.

Prior to implementation, a functional status measurement must posses adequate reliability and validity. Reliability is the consistency of a measurement. Validity pertains to evidence that a test measures what it is intended to measure. Reliability and validity are difficult to differentiate because they are conceptually interdependent.\textsuperscript{7}
A measurement which has reliability but no validity, has no justifiable use or application. Validity, unlike reliability, is neither easily defined nor directly testable. Therefore, validity deserves additional attention.

Validity not only refers to evidence that a test measures what it is supposed to measure, but also indicates the range of inferences that are appropriate when interpreting a measurement, score, or test result. Therefore, the validity of a score attained from a functional ability index is referring to how closely the score reflects the performance of the individual in his or her own environment. Nunnally, a leading author on psychometric theory, states that one validates not a measuring instrument, but rather, some use to which the instrument is put.

The American Educational Research Association defines three basic types of validity: criterion, content, and construct validity. Concurrent and predictive validity are subcategories of criterion related validity. Likewise, empirical and statistical validity are synonyms for criterion validity. Convergent and discriminant validity are component parts of construct validity.

Content validity is defined as whether the items of an instrument adequately represent the domain they are supposed to measure. For example, if a functional assessment instrument were to measure hand function, the items within the scale would have to be sufficiently exhaustive to ensure that several major aspects of hand function were included. Specifically, does the content of the test infer normal hand function?
Concurrent validity is the simultaneous correlation between the variable being measured and the external criterion. Predictive validity is proven when a new test is forecasting the performance of a future test. High predictive validity in a functional assessment is valuable for predicting outcomes based on patient's performance on a particular functional status scale. Michels states that when measurements are taken from two different tests rather than from the same test, the predictive validity of the first test is measured. Concurrent and predictive validity are both component parts of criterion validity.

The criterion validity of a particular test depends on how closely it is related to an external criterion of superior measurement accuracy or the 'gold standard'. The goal of criterion validity is to achieve perfect correlation between the new test and the gold standard. Unfortunately, there is no gold standard for measuring functional ability. No one instrument is perfect at accurately describing any particular individual's functional ability at any one time. Therefore, an alternative form of validity is used. This is known as construct validity.

Construct validity is established by demonstrating its ability to identify the variables it proposes to identify. Functional status measures are frequently compared to a variety of other health and non health related measures believed to be related to a particular type of skill. With construct validity, the goal is not perfect correlation. Instead, the researcher attempts to demonstrate the predicted direction and magnitude of correlation between existing measures and the new test. If the new measure
shows high correlation with the existing measures, the test displays convergent validity.

Convergent validity is only half the requirements for true construct validity. Discriminant validity is also necessary. Discrimination is displayed when a new measure correlates better with a second measure accepted as more closely related to the concept under study than with a third, more distantly related measure. A measure of functional status in arthritis, for example, should show higher concordance with joint range of motion than with erythrocyte sedimentation rate.

Kaufert identifies two approaches to establishing validity of a functional status questionnaire. The first approach involves comparing the final summary of a patient’s functional status attained by the questionnaire with the overall functional classification of the patient by a clinically trained rater (MD, PT, or OT) from observation, examination, or review of clinical records. An alternate approach is to look at the individual items within a functional status questionnaire and make comparisons with the results of clinical examinations or observations. Kaufert further explains that the type of validity evaluated and the design of the validity comparison must ultimately be governed by measurement objectives. Therefore, before any new test can claim to be valid, an operational definition explaining what exactly the test is attempting to measure and the route the test takes to accomplish this task must be clearly set forth. Following will be a look at three common functional status assessments. A general description, the conceptual focus, the measurement dimension, reliability and validity, the mode of administration, and strengths and weaknesses will be discussed.
CHAPTER IV
A COMPARISON OF THREE INDICES

Introduction

Researchers have developed many types of assessments each offering the clinician an added dimension on the well being of their patient. Silverstein et al\textsuperscript{32} states: "Intensifying industry competition and the careful scrutiny of payers, retrospective reviewers, and accreditation agencies now require rehabilitation facilities to prove appropriate resource utilization, program effectiveness, and patient progress toward rehabilitation goals." Some instruments attempt to measure health status or quality of life. These scales are also known as multi-dimensional instruments due to inclusion of social skills. Although these scales purport to measure the broad components of health status, most focus only on two or three dimensions and display very limited information on signs and symptoms or resultant impairments.\textsuperscript{33} The physical function of the individual is the physical therapist's primary interest in the rehabilitation setting. Three common scales of physical function in use today are the Barthel Index (BI), the Katz Index of ADL, and the Functional Status Index (FSI).

Barthel Index

The Barthel Index was developed by Dorthea W. Barthel, PT, and Florence Mahoney, MD.\textsuperscript{9} It was introduced in 1965 to provide a means of measuring the
severity of disability in people whose disease interfered with independent movement of the limbs.\textsuperscript{34}

The Barthel Index consists of 10 areas of physical and self maintenance function. They are feeding, moving from wheelchair to bed and return, personal toileting, bathing self, walking on level surface, ascending and descending stairs, dressing, controlling bowels, and controlling bladder.\textsuperscript{35} Each activity is scored as being 'independent' or 'with help'.\textsuperscript{36} Certain activities are weighted as more significant. For example, independence in continence represents 20 out of a possible 100 points, whereas independence in bathing only maximally represents 5 out of the possible 100 points. The patient is scored in each of the 10 activities and a resultant cumulative score is attained.

A Barthel score of 100 is indicative of sufficient independence in self-care and mobility, eliminating the necessity for attendant care with these basic needs.\textsuperscript{35} In addition, valuable, detailed information may be supplied by the sub scores of each variable.\textsuperscript{35} The Barthel Index is an ordinal scale of measurement, and is therefore only a descriptive label and cannot be used as primary data in any statistical analysis to calculate means or standard deviations.\textsuperscript{36} The Barthel Index is limited to ordinal scale statistical procedures including cross-scale comparisons.

The weighting described in this index has met with considerable criticism. Murdock\textsuperscript{36} states that prioritizing items within a scale is based purely on the experience of clinicians working in rehabilitation and, therefore, is highly subjective. Warner,\textsuperscript{37} in 1976, stated that judgmental weighting is of no benefit and makes no
difference to the value of the test. Gresham,\textsuperscript{38} in 1980, described such weighting as 'frank preferential weighting' and suggested that it 'commends itself as a professionally shrewd prioritization of the most crucial ADL skills'.

The Barthel Index focuses on only one dimension of function; basic self care activities. This is also known as physical self maintenance skills, or basic activities of daily living (BADL).\textsuperscript{36} These are defined as functions necessary for basic survival and include such things as dressing, bathing, walking, stair climbing, and continence.\textsuperscript{5} Instrumental self-maintenance skills, however, are activities necessary to survive in our society and culture.\textsuperscript{36} Examples are shopping and cooking.

Reliability of the Barthel Index is not well documented in the literature because of the large amount of variability possible. How can one guarantee that a patient’s attention or motivation level will remain consistent at two different times? The scale includes an extensive instruction manual to facilitate a standardized assessment. However, according to Murdock,\textsuperscript{36} two shortfalls in the BI’s instructions exist. Although the Barthel Index is one of the few assessments to consider time as an important factor in the completion of an activity, it is not clearly defined. A ‘reasonable time’ leaves the door open to much subjectivity by the assessor. The use of aids and adaptations is also absent. If an individual uses a cane to walk, that person is still granted full credit. Selmen and Barnitt\textsuperscript{40} suggest that the use of any aids or adaptations should accompany a patient’s score. Despite these two shortfalls in instruction, the BI does provide the necessary information for consistent administration.
The inter-tester reliability is high because of the simplicity of the test. The instructions along with this inherent simplicity ensure that two testers will obtain similar results provided the patient performs at similar motivational and attention level. Intra-tester reliability is thought to be of limited value. The functional performance of any subject can change with time due to motivation, fatigue, or actual change in functional skills.

The BI has also shown to possess a fair amount of parallel reliability. The Barthel Index has been extensively compared with the PULSES Profile. The PULSES acronym derives from: Physical condition, Upper limb, Lower limb, Sensory components, Excretory functions, and Support factors. Granger compared the BI with the PULSES Profile and found high test-retest reliability for both scales. Despite the high correlations, there were some noteworthy differences. Admission scores displayed by the PULSES indicated that subjects were more dependent than what the BI showed for the same subjects. Also, the BI is not sensitive to specific functional impairments relating to general health status, communication skills, and psychosocial skills. Granger and Greer also compared the PULSES with the BI and found that Barthel admission and discharge score and PULSES discharge score were closely correlated with discharge outcome in all participating facilities. In general, a Barthel score greater than 20 on admission signified a much greater chance to return home than subjects with a below 20 score. Additionally, patients who attained a discharge score of greater than 60 were much more likely to return home than those with less
than 60. A score of 60 or below is believed to reflect serious limitations in personal care activities.

The BI has been shown to be responsive to changes in function over time. However, some limitations have been identified. A ceiling effect has been noted since a score of 100 (the maximum score) does not equate with normality. A perfect score of 100 does not mean total independence in the home environment. It should be remembered that the original intent for the BI was to provide a gauge for functional improvement in institutionalized elderly with physical disabilities. Therefore, the BI is said to lack responsivity at higher levels of ability. A floor effect has also been identified since a person scoring 0 may either be bed bound but alert or unconscious. Thus, the BI lacks responsivity at extremely low functional levels.

An instrument which demonstrates high reliability, but little validity, is of no use to the clinician. Validity substantiation is difficult with ADL scales because disagreement exists for what constitutes ADL. Law and Lets, when discussing validity as it relates to ADL scales, stated validity is never proven but represents a gradual accumulation of evidence to support the validity.

Construct validity, in light of ADL scales, refers to the parameters of ADL. A scale with construct validity must define precisely what it is measuring. Because Barthel and Mahoney focused solely on basic ADL and provided very detailed operational definitions of each item, the BI is believed to display adequate construct validity.
The BI’s content validity is concerned with the actual items in the test and how they represent the construct being measured. The developers of the BI standardized the index to be used in a hospital environment and not the home environment. Thus, Murdock\textsuperscript{39} argues that the BI displays adequate content validity when used in its intended setting.

Criterion validity is very hard to prove due to the lack of a gold standard for the measurement of function. Often, new tests will be cross compared to existing functional index scales or clinician’s assessment to establish validity. Wylie\textsuperscript{43}, in 1967, compared the BI score with a clinician’s judgment on stroke patients and found relatively high concordance rates across three levels. First, a lower score positively corresponded with a higher death rate. Second, a positive correlation between score increases and independent clinical judgments of improvement were observed. Finally, he found that lower scores corresponded generally with older patients. Until the fundamental difficulties of defining ADL have been tackled, true validity of the BI will continue to present difficulties to researchers wishing to statistically analyze this scale.

The BI is becoming more widely used in the community based setting. Its primary use, however, remains within the specialized rehabilitation setting.\textsuperscript{34} Although the primary mode of administration is direct observation, McGinnis et al\textsuperscript{44} developed a self report version which has increased use after discharge.\textsuperscript{45} Consequently, McGinnis et al\textsuperscript{44} found that scores on the self report version tend to be significantly lower than scores attained through direct observation by a therapist. McGinnis et al\textsuperscript{44} suggest
several possible explanations for this discrepancy. A few include patient anxiety about their limitations, and the therapist's drive for satisfaction from patient progress and success. The BI, in general, is suitable for use with any patient who has physical disabilities provided their cognition and communication are intact. The BI's most common use is with stroke and other hemiparesis patients. However, it is also used for orthopedic and general debility patients.

In summary, the BI demonstrates strengths and weaknesses. Clinically, it is not time consuming. Administration time is approximately 30 minutes. Second, the activities tested are generally representative of the overall functional abilities of physically disabled people. It also provides scores for each activity which allows the clinician and patient to 'see' specific areas of difficulty.

The BI's major limitation is the standardized assessment format. The standardization does not take into consideration personal or contextual differences. The scoring system endpoints are not representative of an absolute maximal functional level. People's abilities can still change even after the end points of the index have been reached.

Katz Index of ADL

Perhaps the most popular scale for assessing physical function is the Katz Index of ADL. This index resulted from many meticulous studies focusing on the behavior of chronic illness in aged persons at the Benjamen Rose Hospital in Cleveland during the late 1950s. The Katz Index, like the Barthel scale, is useful in monitoring individual cases and using aggregate data to document the achievements of entire
rehabilitation programs. However, the Katz Index does have certain unique characteristics placing this index above and beyond the processes of case monitoring and program evaluation. Information from this index, because of its meticulous standardization and wide acceptance, has been beneficial in categorizing the consequences of disease and injury.

Katz, in his original study, found that, after 8 years of attempting to standardize a classification schema of ADL, recovering patients passed through three stages. An early recovery of independence in feeding and continence is followed by an intermediate stage of transfer and toileting. The last and final stage is the regaining of bathing and dressing skills. Katz noticed a striking similarity between the recovery of function from disease or injury and the normal development of children. Katz, therefore, assumed a developmental and hierarchical organization of function in constructing the instrument.

The Katz index consists of six categories of basic ADL: bathing, dressing, toileting, transferring, continence, and feeding. The standardized form for recording ADL evaluations consists of three descriptions of each function. Descriptions range from most to least independent. For example, the first category under the activity of bathing pertains to receiving no assistance. The intermediate descriptor specifies receiving assistance in bathing more than one body part. The final descriptor implies assistance on bathing more than one body part. Katz et al states that two descriptions would permit distinguishing between independent and dependent states;
however, the intermediate state increases observer awareness for subtle distinctions and thereby increases reliability.

After the patient is evaluated in all six activities on the standardized ADL form, the data are converted into an Index of ADL grade. Grades range from A (total independence) to G (total dependence). The grades reflect not only the increasing number of tasks a patient is unable to perform, but also signifies a very specific sequence of loss in ADL independence. An individual’s global letter score indicates an exact pattern of responses to the list of items. A score of B in the Katz Index, for example, means the individual is independent in performing all but one of the six basic ADLs included. Conversely, a score of D indicates the individual is independent in all but bathing, dressing, and one additional function.

The Katz Index is known as a Guttman scale. Guttman scaling is a method of determining whether a set of items (in this case, the six ADL items) form a unidimensional, ordinal score. If they do, knowing an individual’s score indicates the exact pattern of responses across all items.

The theoretical assumption that functional recovery is based on inherent biological and psychosocial phenomena is easily displayed in the Katz scaling. For example, someone classified in grade G (completely dependent) would be expected to regain independence first in feeding or continence, then transfer and so on. Such functional recovery parallels the normal development of children.

Labi and Gresham raise three important issues when discussing the sequence of return of functional abilities as it relates to development in children. First,
environmental adaptation will certainly affect an individual’s ability to function independently even though there is no change in the individual’s ability. Second, is earlier return of one function over a different function reflective of that function’s biological primacy over another? Could this be due, rather, to priorities and emphasis practiced in a specific rehabilitation setting? Third, does function always return in sequence or in clusters?

Despite the debatable theoretical assumptions of the Katz Index, a wide array of uses are identified in the literature. The index has led to objective information about the progression and regression of chronic illness and has contributed new information to the body of knowledge concerning functional status. Sidney Katz, MD, the developer of the index, has used it many times in his own research. Katz, in 1967, found patients with fractured hip experience most full and partial recoveries by one year post fracture with negligible recovery after two years. Similar recovery period was noted following stroke rehabilitation; most significant recovery occurred within six months with only negligible recovery after two years. Many other examples of Katz index utilization have been identified in the literature. These examples demonstrate how the index can attempt to answer questions concerned with recovery time, target populations for rehabilitation services, and the amount of assistance needed to maintain community independence.

Reliability of the Katz Index is quite high. Agreement ratios of .68 to .98 were attained between different professional raters. Test-retest reliability for respondent self-reports produced coefficients from .61 to .78. The degree of discordance
ranged from 0% to 24% across observers and from 0% to 10% over time when a
second interview was completed three hours after the first.15

Validity studies of the Katz Index of ADL are very limited. This is again due
to the lack of a 'gold standard' for comparison. The index has, however,
demonstrated the ability to successfully predict the need for future assistance following
discharge.15 For example, it was found that among hemiplegic patients, 79% in grades
D, E, F, or G at discharge were receiving nonfamily attendant care (e.g., nursing
home) one year after stroke. Conversely, only 45% of the hemiplegic patients in
grade B or C at discharge were receiving nonfamily attendant care. Katz15 believes
this predictive information is especially important in decision making about how and
when, in the course of illness or the aging process, application of preventive and
restorative resources is most effective.

In summary, the Katz Index has its strengths and weaknesses. Strengths
include a practical utility in achieving broad functional classifications, widespread
acceptability, and brevity of administration. Weaknesses include debatable theoretical
validity, and its lack of exhaustive categories. For example, there is no category for
patients who use both a device and human assistance.5 This may reduce the reliability
of the instrument in some populations. A major weakness is its failure to include an
item on ambulation.33

Functional Status Index

A third common functional assessment is the Functional Status Index (FSI).
The FSI defines function as three distinct but related dimensions: the degree of
dependence, the degree of difficulty, and the amount of pain experienced in performing specific activities of daily living.\textsuperscript{47} The FSI, unlike the previously described scales, measures both basic and instrumental ADL.\textsuperscript{5} Activities assessed include Mobility, Personal Care, Home Chores, Hand Activities, Vocational, and Avocational activities. Functional dependence is assessed on a five point scale, where one equates to complete independence, two signifies use of equipment, three equals use of human assistance, four identifies both use of equipment and human assistance, and five equals complete dependence.\textsuperscript{5} Pain and difficulty are measured on a seven point ordinal scale with one signifying no pain/difficulty and seven representing severe pain/difficulty in completing a particular activity. The original index measured performance across 45 specific ADL. This version took a trained interviewer 1 to 1.5 hours to complete, making it too long for regular use in the clinic.\textsuperscript{47} Exploratory factor analysis reduced the original 45 down to a more feasible 18 specific ADL.\textsuperscript{5} Administration time was, therefore, reduced to approximately 30 minutes. Data are attained either by patient self report or by one-on-one interview format.\textsuperscript{5}

The developers of the FSI were interested in formalizing a measuring device which could attain reasonable levels of reliability and validity using readily accessible sources of information.\textsuperscript{47} Second, the index should be brief to facilitate use in clinical investigations while the level of sensitivity falls somewhere between the crude measures, such as the American Rheumatism Association's (ARA's) Functional Classification device, and the exhaustive and time consuming 100+ ADL assessment
instruments used in many rehabilitation centers. Alan M. Jette, one of the originators of the FSI, has done significant research on the reliability and sensitivity of this scale.

Internal consistency estimates the reliability based on intercorrelations among items within the same functional category. Using the Spearman Brown Formula, the FSI demonstrated correlations from .66 to .91 for all factors. Each functional category, therefore, demonstrates adequate to high representation levels within each functional category. For example, activities of walking inside, stair climbing, and chair transfers, as a group, adequately represent the area labeled Gross Mobility.

A critical problem in assessing degree of test-retest and interobserver reliability is selecting an appropriate interval of time between administration of the index. The interval needs to be long enough to minimize the potential confounding influence of memory but needs to be short enough to minimize the possibility of real changes in the function being assessed. Jette chose an interval of one to three days. An interclass correlation coefficient was used to calculate test retest and interobserver reliability. Generally, individual measures of functional dependence achieve test-retest and interobserver reliabilities of .6 or above. On average, Jette demonstrated that inter-observer reliability of the measures equals and sometimes exceeds the level of test-retest reliability. Activities where interobserver exceeded test-retest reliability were degree of dependence in visiting family and friends, attending meetings, and driving a car. Clinical implications of this finding are encouraging to clinicians who work in settings where different interviewers would be employed. A few areas
where test retest reliability was substantially greater than the inter observer reliability was in dependence level while vacuuming a rug and putting on a shirt.

Other limitations in reliability were also pointed out by Jette. For example, inter-personal activities (Avocational Activities) achieved lower reliability levels as compared to the other five functional categories. Additionally, the degree of dependence in opening containers displayed rather low correlations for both types of reliability.

All previously described data come from a study focusing exclusively on one particular patient population, adult rheumatoid arthritis patients. A similar study completed by Jette and Deniston, focused on a more diversified patient population, including the chronically ill, in addition to arthritic patients. Inter-observer reliabilities ranged from .61 to .78 which again demonstrates adequate to good reliability across different patient populations. The main use of the FSI clinically, however, remains with arthritic patients and should be used with caution when assessing other patients.

Validity of the FSI has been largely assessed through cross comparisons with ARA stage in addition to physicians’ estimates of disease activity and functional ability (i.e., construct validity). Jean T. Shope, in his 1983 study, demonstrated the FSI’s ability to correlate positively and significantly with both professional opinion and ARA functional classification thereby proving the instrument’s construct validity. However, Deniston and Jette, argue that FSI scores demonstrate stronger correlations with client reports rather than professional assessments of the client’s joint condition. Thus, it appears that client and professional definitions of joint condition differ and
only the client's definition relates to the score on the FSI. Future research is necessary to clarify validation and definition concerns when using the FSI.

In conclusion, the FSI has many benefits. Guccione et al states that when considering ambulation the FSI is one of the few assessments that can distinguish between human and mechanical assistance. Also, by quantifying three dimensions of function (i.e., dependence, difficulty, and pain), the FSI attempts to incorporate greater sensitivity into its design. The FSI assesses both instrumental ADL and social interaction in addition to basic ADL.

Limitations of the FSI are few. The age of the index may, in itself, be its biggest weakness. Despite the extensive reliability and validity studies, it remains to be less established than either the Barthel or the Katz Indices. Also, the index should be used with caution with non arthritic patients because of limited use with a diversified population. Another weakness is its lack of sensitivity for detecting change when patients have previously modified their ADL to meet their particular needs.

Summary of Physical Function Tests

Each functional ADL scale has its own unique advantages and clinical applications. Likewise, each has its own disadvantages and shortfalls. This brief critical review of three common functional ADL scales reveals that no one instrument fulfills all five criteria for sound measurement ability which were identified by Liang and Jette in chapter III. The functional indices discussed, however, do show great progress in establishing adequate scientific rigor when compared to earlier efforts of classification.
Comparisons between each index has been cited in the literature. Gresham, Phillips, and Labi\textsuperscript{38} identified 11 most common ADL variables used in ADL scales written in the English language. The Barthel index evaluates all 11 variables whereas the Katz Index evaluates only seven variables. Guccione, Cullen, and O’Sullivan,\textsuperscript{33} in a similar comparison, found the Katz Index to measure only five of 12 common basic ADLs. The FSI and Barthel Index were found to measure six and eight of the 12 ADL respectively. The FSI, however, also measured five of eight common instrumental ADLs, whereas both the Barthel and Katz Indices measured none of the identified instrumental ADLs.

Gresham, Phillips, and Labi\textsuperscript{38} found that the Katz Index and Barthel Index show significant agreement in their ability to classify patients in the same category. This finding suggests both indices possess similar sensitivity levels. It is also evidence for construct validity (i.e., convergent validity). Donaldson, Wagner, and Gresham,\textsuperscript{51} however, found that the Barthel Index reflects changes in physical function which will not necessarily move the patient into the next Katz category. The FSI, when compared to the Barthel Index, is more sensitive for discerning human verses mechanical assistance when evaluating ambulation ability.\textsuperscript{33} The Katz Index, as mentioned previously, does not include ambulation ability as part of its index.

Each index lends itself to adequate clinical feasibility. Each may be completed in much less than one hour if done by a trained observer.\textsuperscript{38,47} However, in regard to actual clinical use, painstaking orientation, training, and onsite demonstration trials must be completed for each index to ensure reliability and validity.\textsuperscript{38} The FSI, unlike
the Barthel and Katz indices, has been specifically tested for use in non-institutionalized individuals with arthritis, thereby increasing its overall use\textsuperscript{15}. It should be noted, however, that although the FSI possesses better scientific rigor than either the Barthel or the Katz, it has experienced minimal acceptability because of its short existence relative to the Barthel and Katz Indices. The latter two indices have been around for nearly 30 years. Actual clinical usage of both the Barthel Index and the Katz index are observed in survival studies on recovered stroke patients. Likewise, the FSI has been implemented extensively in similar studies involving arthritic individuals.
CHAPTER V
CONCLUSION

Clinicians are now focusing primarily on functional outcomes rather than objective tests. However, disagreement exists concerning standard definitions of terms like rehabilitation, health status, and functional status. Function, for example, concerns itself with many confounding variables rendering a standard definition virtually impossible. Researchers conveniently define function to be compatible with the patient type and/or the context in which medical care is given. Function in a specific sense implies performance of a body part (function of the shoulder). Function in a holistic sense describes the performance of an individual in activities of daily living. Consequently, such shrewd definitional selectivity leads to measurement problems.

The ICIDH has attempted to provide a conceptual framework for the measurement of function. Health is separated into physical manifestations, symptoms of feeling, and functional status. Functional status is further separated into physical, emotional, mental, and social function. Standardization of these terms is exemplary of what is needed for other closely related yet questionable terms.

Likewise, the consequences of disease have been defined. Any particular disease results in an impairment, disability, and possibly, a handicap. However,
disagreement still remains when discussing the placement of functional limitations within the consequences of a disease. The ICIDH states functional limitations fit under the category of impairments. Other researchers believe functional limitations should be included under the category of disability. Research leading to standardized definitions must be continued if successful measurement of function is possible.

The purposes of functional status tests are to plan treatment, determine treatment effectiveness and efficiency, maintain continuity of care, develop treatment resources, and improve treatment resources. Unfortunately, a universal, all encompassing assessment measure to document functional outcomes has not been developed. Most importantly, no instrument has demonstrated adequate reliability and validity. Immeasurable costs and/or benefits, individual differences, and the absence of human clinical judgment are additional reasons for measurement difficulty in functional assessments. Furthermore, assessment of an individual’s quality of life has moral implications. One’s quality of life depends on individual perceptions. Therefore, the inherent de-humanizing effect of functional assessments must be acknowledged, thus preserving the individual attention that patients expect.

Functional scale implementation difficulty pose yet another obstacle preventing the growth and development of functional assessments. Functional status measures must be introduced directly into current medical school curricula where young medical professionals’ attitudes towards new innovations are shaped.

However, before implementation, a scale must demonstrate adequate reliability and validity. A measurement with reliability but no validity has no justifiable use or
clinical application. Presently, functional assessments lack a superior external criterion necessary for criterion validation. Therefore, functional assessments rely on construct validation rather than criterion related validation.

A detailed analysis of three common functional assessments revealed strengths and weaknesses of each. The Barthel Index appears to cover basic ADL activities the best. However, the FSI, unlike both the BI and Katz, does include some instrumental ADL activities. The Barthel Index is also quite sensitive to real functional change. The BI's major limitation is the standardized assessment format which does not consider personal differences or contextual differences.

The Katz Index has been used to define time frames for progression and regression of certain conditions. The absence of ambulation activities is the major limitation of the Katz Index. Validity studies of the Katz Index is also very limited.

The FSI, because of its interval scale, has been extensively studied for sound reliability and validity. The FSI is also more sensitive than the Barthel Index for evaluating ambulation. The limited existence of the FSI is probably its greatest limitation. Clinical exposure must be enhanced before the FSI can be compared with the more established Barthel and Katz Indices.

In conclusion, more research is necessary in the area of functional status measurement devices. As clinicians and insurance companies demand more functional outcomes, a valid and reliable measurement tool is necessary. Based on this literature review, important areas for future study parallel areas cited be Liang and Jette.\(^1\)
Most importantly, validation of existing and new functional status instruments is necessary. Improved external criteria such as computerized gait analysis may provide a way to validate clinical scales of activity. Also, comparing new scales against timed performance standardized for age and gender might also be used for external criteria.

Second, the refinement of the operational definition of function is necessary. Functional status measures must focus only on measuring function and not on the reason for the dysfunction. Attempts to ascribe a cause for dysfunction leads to bias either from the patient or from the rater.

Third, disease specific measures of musculo-skeletal function appears to be a way for increasing sensitivity to a level which allows for assessing and monitoring the changing functional status of an individual.

Finally, the convenience of administration must be further improved if widespread use is to be expected. Functional status instruments can be shortened by eliminating redundant items through factor analysis and similar statistical techniques.

Presently, a great area of concern to all medical providers is health care reform. Throughout the eighties, health care costs have sky-rocketed leaving many individuals in the United States without health insurance. Many blame these high costs on outrageous medical provider fees. However, only a small amount of money actually goes towards medical costs. Missed work days accounts for the largest cost of the injury.³
Health insurance companies are increasingly looking for functional outcomes in the documentation of rehabilitation clinicians. Functional assessments may be thought of as a snapshot displaying one's functional ability at a particular moment in time. Third party payors use these snapshots to determine the possibility for further improvement in the patient's condition. Documentation for reimbursement must objectively and quantifiably show third party payors that further improvement is likely. A multitude of functional assessments known as Functional Capacity Evaluations are now being utilized in this manner.

The challenge of assessing physical function is still present today as it was 30 years ago. Steinbacker,\textsuperscript{15} back in 1949, stated: "the evaluation of therapeutic results has long been confused by the failure to distinguish between the effects of therapy on rheumatic activity in contradistinction to its influence on functional capacity."

It is hoped that this literature review has made the practicing clinician more cognizant about selecting a functional status index. There are many variables affecting the efficacy of a particular assessment in the clinician's specified setting. Awareness of these variables will lead to increased efficacy of functional status assessment which is hoped to ultimately facilitate improved patient care in the area of rehabilitation medicine.
REFERENCES


3. Rud A. Personal Interview at the Medical Center Rehabilitation Hospital; December 28, 1993; Grand Forks, ND.


