

# Choosing appropriate measures in inpatient rehabilitation

Anne Sinnott, Sarah Dean

**This article aims to assist inpatient rehabilitation clinicians to choose appropriate measures. Emphasis is given to measurement processes and interpretation. The authors provide examples of measures commonly reported for five rehabilitation diagnostic groups: stroke, traumatic brain injury, spinal cord injury, multiple sclerosis and cardio-respiratory rehabilitation.**

Given the need for accountability, increasing competition for scarce resources and the demands of busy inpatient rehabilitation environments, choosing the most appropriate measures is crucial. Measurement of outcome means simply the measurement of consequences for the individual (Wade, 1992). Measurement is carried out for a number of reasons: to diagnose, to measure severity or extent of a feature present in an individual, to make a prognosis, to evaluate change over time and to measure outcome. These can be broadly split into three categories of measurement: diagnostic, predictive and evaluative (outcome). All three may be required in the inpatient setting.

Without some sort of framework for interpretation of outcomes, measurement may be inadequate. This is particularly so if measurement is limited to the consideration of activities of daily living (ADL) without exploration of individual roles, expectations and perception of opportunities. Partly to address this issue, the World Health Organization (WHO) published a framework in 1980, the International Classification of Impairment, Disability and Handicap (ICIDH) (WHO, 1980) that was designed to represent the 'consequences of disease'. This was subsequently refined to the International Classification of Functioning, Disability and Health (ICF) (WHO, 2001). The ICF is discussed in detail by McPherson et al (2005).

### OUTCOME MEASUREMENT

This article focuses on 'outcome measurement' for inpatient settings. While it is acknowledged that measures of outcome are only tools and not necessarily capable of improving clinical practice, standardized outcome measures can be used to assist clinical reasoning processes to:

- Determine the strengths and weaknesses of the patient
- Help determine, in a systematic way, whether a patient is ready to move to the next level in rehabilitation or even discharge from inpatient care
- Following discharge, to track the patient's functional independence, participation levels in society, general health status and health-related quality of life (HRQoL) over time.

Most significantly, from a patient-care perspective, outcome measures can be utilized to help interdisciplinary teams communicate effectively. They allow this to occur in a standardized way particularly with regards to interventions needed for patients. From an audit perspective, outcome measures may help determine strengths and possible weaknesses of a service. This may enable a health-care provider to justify funding and resource allocation. In addition, standardized outcome measures allow research to be undertaken that may advance an area of knowledge in rehabilitation. For example, 'outcomes' or consequences of rehabilitation interventions in one inpatient stroke unit could be compared internationally with outcomes in other stroke units. Information can be shared and compared in a standardized format. However, it is rare that simple and international benchmarking occurs without problems.

### OTHER CONSIDERATIONS

There has been interest in recent years in general health status. This can create a clinically significant problem in inpatient rehabilitation settings owing to the complexities of health status or well-being in the presence of co-morbidity. For example, the validity and/or applicability of

Ms Anne Sinnott is Lecturer (Rehabilitation) and Dr Sarah Dean is Senior Lecturer (Rehabilitation), Rehabilitation Teaching and Research Unit, Wellington School of Medicine and Health Sciences, University of Otago, PO Box 7343, Wellington, New Zealand

Correspondence to: Ms A Sinnott

health-status measures in persons with cognitive impairment are largely unknown. It has been suggested therefore that these measures are interpreted with caution in individuals with cognitive impairment (Riemsma et al, 2001). One published measure, the Multiple Sclerosis Impact Scale (MSIS-29), is described as a patient-based, disease-specific measure of the health impact of multiple sclerosis. This measure shows promise and authors report that all psychometric criteria were satisfied and that preliminary evidence of responsiveness was demonstrated (Hobart et al, 2004).

Measurement of HRQoL is topical but can often be problematic. While the problems associated with HRQoL are beyond the scope of this article, it has been suggested that one issue of interest in rehabilitation is the lack of consensus over which measure to use. A Cochrane review of exercise-based rehabilitation for coronary heart disease found 18 different instruments were used to assess HRQoL with eleven studies reporting it as an outcome (Joliffe et al, 2000). However, efforts have been made to reach a consensus on the most appropriate measures to use for inpatient rehabilitation of traumatic brain injury (Bullinger, 2002) and spinal cord injury (SCI) (Wood-Dauphinée et al, 2002).

### **CHARACTERISTICS OF MEASUREMENT**

Ten important considerations that should determine choice of measures have been identified and are listed below.

#### **Purpose and utility**

It is important to consider the following questions to define the purpose and use of a standardized measure:

- Does it cover information the clinician is interested in?
- To whom does the information make sense? Is it being used to communicate with a narrow or wide audience; how complex is the measure required to be?
- Is a detailed description or a simple classification required?
- How long will it take to complete?
- How much will it cost?
- How will patients and clients respond to format and content e.g. the ethics of the questions? Are the questions appropriate in the culture or are they distressing?

A number of other issues fall into this category depending on the specific purpose. For example, the real-life relevance of measurement cannot be emphasized enough (Whiteneck, 2002).

#### **What needs to be measured?**

If a measure of anxiety is only used, then this will not be particularly effective in determining whether a treatment for depression is effective. This is a simplistic but not unknown example of what happens when setting out to measure something without fully defining what is that is being addressed. For example, the 36-question health survey (SF-36) has a mental-health component that measures both anxiety and depression as part of the whole mental health score, but it does not differentiate between anxiety and depression.

#### **Why is a measure required?**

It is necessary to clarify why a measure is needed because there are a number of parties interested in rehabilitation outcomes. It is important to question whose perspective the measure takes. If the acceptability of the service to clients is a primary interest, it may be preferable to choose a measure that focuses on satisfaction with services. However, if the desire is to address acceptability of the service from a purchaser's point of view, a satisfaction questionnaire may not be adequate.

#### **Is it important?**

It is necessary to know whether something about the importance and value of the health condition is being addressed. It may also be relevant to determine from whose perspective it is important. There are many measures available that may address a particular question very well; however, it is important to assess whether the question being asked will result in quality information that can be used effectively.

#### **Is it ethical?**

There should be a concern about how ethical it is to ask specific questions. It is usual to question the ethics of a particular research study, but it is also important to consider the ethics of the questions used to measure outcome in non-research situations.

#### **Does it address what it is hoped will change?**

This is an important question when looking to evaluate a therapy or management strategy. The sole use of a measure of ADL would not give the desired results when hoping to document improved perceived health. This is sometimes seen expressed in terms of an intervention failing because change in the measure has not been noted. If an inappropriate measure was chosen, then this would be the result, and so the interpretation that the intervention was ineffective is somewhat worrying.

### How widely used is this measure?

The use of a measure and whether it is used often or seldom is important to consider. If a measure was first described in 1980 and has not been used since, there may be good reasons why. In terms of international benchmarking, the more widely used a measure, the more likely it is that there will be an understanding of the subject. Given that measurement is about communicating information, this is not an inconsequential aspect of choosing a measure.

### Is it usable?

One of the most widely used questionnaires about health has been the Sickness Impact Profile. However, it is very long with 136 questions and this often means that, although in other ways it may meet the clinicians needs, it is unusable. For this reason the SF-36 is often used. There are now versions of the SF-36 with only 12 or 8 items (see SF-36.org). Another problem occurs when members of a team are not available to complete a measure designed for team completion.

It is necessary to ask whether the measure has a good instruction manual indicating how to use it correctly, and whether it costs a lot to purchase? For example the Functional Independence Measure requires purchase and specific training that may or may not be easily accessed.

### Focus, scope and level of measurement

Focus and scope are sometimes used interchangeably but it is probably less confusing, and more accurate, to try to define them in the following way. Focus concerns what the measure is directed towards getting information about. This is where the ICF is useful. Is body-structure or function, activity or participation, or something else required? In contrast, the scope of the measure means how broad or extensive the measure might be. Level most commonly refers to the mathematical properties of the measure. Does it add up and, if so, how? These levels include nominal, ordinal, interval and ratio data and there are distinct characteristics that differentiate these levels. The main effect is how to interpret scores or numbers used in the measure. The more complex the level of the measure, the more powerful the statistics that can be used. Using more robust statistics can be helpful in research studies and allows results to be generalized.

### Psychometric properties and statistics

Psychometric properties and statistics are questions most people tend to think of immediately when asked to evaluate a measure. While complex, and well beyond the scope of this article to discuss in detail, there are a number of characteristics that fall into this category. The most important ones to think about first are outlined

**TABLE 1.**  
**Basic psychometric considerations**

Validity	Does it measure what it sets out to measure? Face validity indicates that a measure tests what it purports to measure Content validity refers to the relevance of the items in the measure and how they relate to the theoretical domain of content – most useful with questionnaires Criterion-related validity is the most practical approach to validity testing and the most objective; and for that reason this term is used widely in articles which describe the development of new measurement tools. It involves the comparison of results obtained using the new measure with those obtained using a tried and tested for 'gold standard' measure. Construct validity establishes the ability of an instrument to measure an abstract construct and the degree to which the measure reflects the theoretical components of the construct.
Reliability	Is the measure going to give the same result – can the score be trusted? Again this can be tested in different ways, the most common being: Inter-rater reliability = reproducibility by different raters Intra-rater reliability = reproducibility by same rater at different times
Responsiveness	The measure's ability to detect real change (sometimes called sensitivity but the more accurate term is responsiveness) The more items (or gradations within items) – usually the more responsive
Sensitivity	A measure of validity of a screening procedure, the probability that someone with a disease or condition will test positive.
Specificity	A measure of validity of a screening procedure. The probability that someone who has the disease or condition will test negative

in Table 1. In addition, it is important to ascertain normative values, i.e. what is an expected normal score for an individual or group of individuals given the set of circumstances.

One area that causes confusion is dealing with the mathematical part of evaluating measures or statistics. Essentially, there are two basic reasons for using statistics. The first is to allow data to be described to provide general observations (referred to as descriptive statistics). The second allows conclusions or comparisons to be made from the study sample to populations (generalization) (Peat, 2001).

### MEASUREMENT CHOICES FOR INPATIENT REHABILITATION

No measure is perfect. It is a matter of weighing up the strengths and weaknesses. The choices

available are exhaustive. However, a measure used for an individual might be totally inappropriate for a large sample. For example, when wishing to know about the functional status of thousands of people with traumatic brain injury, the functional independence measure (FIM) would not be the measure of choice. This is because the FIM is an observation-based measure, which means that the 'rater' must see the person carry out the tasks before he/she can be assigned a certain level of function.

It is important to consider that there is often a trade-off between sensitivity and specificity and reliability and responsiveness. For example, despite widespread use of the FIM in acute SCI rehabilitation, critics suggest it is not sensitive enough to detect changes in function that may occur. In this situation the FIM's motor

**TABLE 2.**  
Examples of measures used for five diagnostic groups by the International Classification of Functioning (ICF)

ICF Classification	Stroke	Traumatic brain injury	Spinal cord injury	MS	Cardio-respiratory
Body Structure and Function	Length of stay Ashworth Scale EMG Studies Pain Scales ROM Sensation Berg Balance Scale Blood pressure Beck Depression Scale	Length of stay Glasgow Coma Scale Blood pressure HADS Beck Depression Scale	Length of stay MMT EMG Studies Ashworth Scale Pain Scales ROM Sensation HADS Beck Depression Scale	Ashworth Scale EMG Studies Pain Scales ROM Sensation Berg Balance Scale Beck Depression Scale	Mortality Cholesterol Blood pressure Chest pain Shortness of breath Peak oxygen uptake Borg scale (breathing perception) St George's Respiratory Questionnaire HADS
Activity	Barthel ADL Index FIM MAS	Barthel FIM DRS	FIM SCIM QIF ADL abilities	Barthel FIM MAS DRS CIQ	Smoking Cessation 6-minute walk test Exercise capacity test General practitioner follow-up
Participation	COPM	COPM CHIEF	COPM CHART CHIEF CIQ	CHIEF IPA	Chronic respiratory disease questionnaire Attendance at rehabilitation

ADL = Activities of daily life, CHART = Craig handicap assessment rating technique, CHIEF = Craig Hospital inventory of environmental factors, CIQ = Community integration questionnaire, COPM = Canadian occupational performance measure, DRS = Disability rating scale, EMG = Electromyography, FIM = Functional independence measure, HADS = Hospital anxiety and depression scale, IAP = Impact on participation and autonomy questionnaire, ICF = International Classification of Functioning, Disability and Health, MAS = Motor assessment scale, MMT = Manual muscle testing, MS = Multiple sclerosis, QIF = Quadriplegia index of function, ROM = Range of motion, SCIM = Spinal cord independence measure. From World Health Organization (2001)

**TABLE 3.**  
Measures identified for five diagnostic groups for Health Related Quality of Life and Health Status

ICF Classification	Stroke	TBI	SCI	Multiple sclerosis	Cardio-respiratory
Health Related Quality of Life (HRQoL)	EUROQoL WHOQoL	EUROQoL WHOQoL		EUROQoL WHOQoL Multiple Sclerosis Quality of Life Index (MS-QLI)	Chronic respiratory disease questionnaire
Health Status	SF-36 SIP Nottingham Health Profile		SF-36 SIP	Multiple Sclerosis Impact Scale (MSIS-29)	SIP Nottingham Health Profile

ICF = International Classification of Functioning, Disability and Health, WHOQoL = World Health Organization Quality of Life, SF-36 = 36-item Health Survey, SCI = Spinal cord injury, SIP = Sickness impact profile, TBI = Traumatic brain injury

items are too difficult for individuals with tetraplegia, while the cognitive items measure aspects that are not affected by the spinal cord lesion (van Tuijl et al, 2002). Therefore, while the FIM may be effective as a disability measure and useful in evaluating the effects of rehabilitation on a program analysis basis, it does not provide the detailed information that is useful in the evaluation of ADL performance following acute SCI.

## RECOMMENDED MEASURES FOR CONSIDERATION

For consideration of the five diagnostic groups mentioned, relevant articles were identified from Medline and CINAHL and PSYCHINFO databases for the period 1996–2004, cross-indexing each with ‘inpatient’, ‘measurement’, ‘outcome measures’, ‘consensus’ and ‘rehabilitation’. This search was augmented through articles identified in reference lists and from work undertaken by one author specifically in the field of outcome measurement for inpatient SCI. *Table 2* lists examples of measures that are categorized according to the ICF domains. *Table 3* lists examples of measures of HRQoL and general health or perceived health status. While not specifically endorsed by these authors, all were reported as useful in inpatient settings.

## CONCLUSIONS

The use of standardized outcome measures has the potential to allow all members of an interdisciplinary team in inpatient rehabilitation environments to make a clear distinction between three things: the process of assessment (requiring interpretation); measurement (quantification of an observation against a standard), and a determination of the consequences of the health episode and rehabilitation process for the individual experiencing the health problem.

To interpret information from measures requires an understanding of measurement itself and the context in which it has been used. With use of the ICF framework, inpatient rehabilitation clinicians are well placed to recognize the strengths and weaknesses of measurement and to agree on standardized outcome measures that would improve

the understanding of consequences over time in terms of a person’s level of function. **HM**

*Conflict of interest: none*

- Bullinger M, Azouvi P, Brokks N et al (2002) Quality of life in patients with traumatic brain injury – basic issues, assessment and recommendations. *Restor Neuro Neurosci* **20**(3-4): 111–24
- Hobart JC, Riazi A, Lamping DL, Fitzpatrick R, Thompson AJ (2004) Improving the evaluation of therapeutic interventions in multiple sclerosis. *Health Technol Assess* **8**(9): 1–48
- Joliffe JA, Rees K, Taylor RS, Thompson D, Oldridge N, Ebrahim S (2001) Exercise-based rehabilitation for coronary heart disease. *Cochrane Database Syst Rev* 2001(1) update, software, Oxford
- McPherson KM, Levack W, Kersten P (2005) A new classification for outcomes in illness and injury. *Hosp Med* **66**(4): 210–14
- Peat J (2001) *Health Science Research – A Handbook of Quantitative Methods*. Allen & Unwin, Crows Nest, New South Wales
- Riemsma RP, Forbes CA, Glanville JM, Eastwood AJ, Kleijnen J (2001) General health status measures for people with cognitive impairment: learning disability and acquired brain injury. *Health Technol Assess* **5**(6): 1–100
- van Tuijl JH, Janssen-Potten YJM, Seelen HAM (2002) Evaluation of upper extremity motor function tests in tetraplegics. *Spinal Cord* **40**: 51–64
- Wade D (1992) *Measurement in Neurological Rehabilitation*. Oxford University Press, UK
- Ware JJ, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* **30**: 473–83
- Whiteneck GG (1994) Measuring what matters: key rehabilitation outcomes. *Arc Phys Med Rehabil* **75**: 1073–76
- Wood-Dauphinée S, Exner G, the SCI Consensus Group (2002) Quality of life in patients with spinal cord injury – basic issues, assessment, and recommendations. *Restor Neuro Neurosci* **20**: 135–49
- World Health Organization (1980) International Classification of Impairment, Disability and Handicap (ICIDH). WHO, Geneva
- World Health Organization (2001) International Classification of Functioning, Disability and Health (ICF). WHO, Geneva

## KEY POINTS

- Outcome measures are useful for inpatient assessment, the measurement of a rehabilitation intervention and for evaluating the consequences for a person receiving rehabilitation.
- Interpreting outcome measures requires an understanding of the measurement itself and the context in which it is being used.
- The International Classification of Functioning, Disability and Health provides an excellent framework for recognizing the strengths and weaknesses of a measure.
- Use of standardized outcome measures can facilitate interdisciplinary rehabilitation within specific inpatient settings, between comparable settings as well as internationally.