

Outcome Measures: Glossary

International Classification of Functioning, Disability and Health (ICF)	<p>A classification and model developed by the World Health Organization in 2001. It includes the following concepts: body functions and structures; activity and participation; personal factors and environmental factors. For more information, consult the World Health Organization’s ICF website: http://www.who.int/classifications/icf/en/.</p>
Interpretability	<p>The degree to which one can assign qualitative meaning (that is, clinical or commonly understood connotations) to an instrument’s quantitative scores or change in scores (<u>Fitzpatrick et al. 1998, MOT 2002</u>).</p> <p>Common measures include the <i>SEM, MDC, and MCID</i> (expanded on below). These are termed “clinically relevant values” because these values allow a clinician to determine when a patient’s change in score actually indicates a change in the patient’s health/ability/functional status.</p>
Standard Error of Measurement (SEM)	<p>Amount of error that can be considered measurement error. In a set of repeated scores, response consistency is measured (i.e. test-retest reliability or intra-rater reliability) and measurement error can be determined. (<u>rehabmeasures</u>)</p> <p>When data is available in articles to calculate SEM, we did so using the following formula:</p> $SEM = \text{standard dev. of sample mean} \times \sqrt{1 - ICC}$ <p>where:</p> <ul style="list-style-type: none"> ▪ the standard deviation of the sample mean is for the first trial, if multiple trials were conducted ▪ ICC is the intra-class correlation for test-retest reliability or intra-rater reliability
Minimal Detectable Change (MDC)	<p>Refers to the minimal amount of change in the instrument’s score that reflects true change (noticeable change in ability) by a patient between two time points (ensures change isn’t the result of measurement error). (<u>Strokengine</u>)</p> <p>When data is available in articles to calculate MCID, we did so using the following formula:</p> $MDC = SEM * 1.96 * \sqrt{2}$
Minimally Clinically Important Difference (MCID)	<p>The smallest difference in score in the domain of interest which patients perceive as beneficial. Changes between baseline and follow-up are examined in relation to their benchmark for a MCID, which was the patient’s follow-up assessment in a transition item of whether they were worse, better, or the same compared with the baseline assessment. (<u>Strokengine</u>). This value is less commonly reported, as it requires asking patients to assess what change in score provides actual tangible improvements in the underlying construct being measured.</p>

Outcome Measures: Glossary

Reliability	Does the instrument produce results that are reproducible (free from random error) and internally consistent? (expanded on below) (<u>Fitzpatrick et al. 1998, MOT 2002</u>)
Internal Consistency	<p>Based on the basic principle that several related observations provide more reliable estimate than a single observation. Therefore, it measures how inter-correlated items of a measure are with each other and the total score; that is, if the items are all measuring aspects of a single attribute or construct.</p> <p>Generally assessed using Cronbach's alpha, item-to-item and item-to-scale correlations. (<u>Fitzpatrick et al. 1998, Terwee et al. 2007</u>)</p>
Reproducibility	<p>Does the score give us an actual indication of the individual's status or is it noise/ random error? We can measure this by taking repeated measures of an individual and examining whether the results stay the same, when the domain of measurement has not changed for the individual. The length of time between measurements should be long enough that individuals do not recall their previous answers, but not so long that it is possible the domain being measured has changed.</p> <p>Generally assessed using correlation statistics including intra-class correlation coefficient (ICC), Pearson's coefficient, Spearman's coefficients and kappa coefficients. (<u>Fitzpatrick et al. 1998, Salter et al. 2005</u>)</p>
Test-retest reliability	Repeat measurements of the same test are taken by an individual and their scores are examined for degree of difference between the tests
Inter-rater reliability	Repeat measurements of the same individuals are made by <i>different</i> raters and the scores are examined for degree of difference.
Intra-rater reliability	Repeat measurements of the same individuals are made by the <i>same</i> rater over a period of time and scores are examined for degree of difference
Validity	<p>Does the instrument measure what it claims to measure? (<u>Fitzpatrick et al. 1998, MOT 2002</u>)</p> <p>The degree to which scores of an instrument are consistent with hypotheses, with regard to: 1) internal relationships, 2) relationships to scores of other instruments, and 3) differences between relevant groups.</p>

Outcome Measures: Glossary

Responsiveness	Does the instrument detect clinically important changes (perhaps reflecting therapeutic effects) over time? (Fitzpatrick et al. 1998) Generally assessed with effect size or standardized response mean (expanded on below)
Effect size	The size of change on a measure that occurs to a group between assessments compared with the variability of scores of that measure. Commonly calculated as the difference between mean scores at assessments divided by the standard deviation of baseline scores. Usually expressed in standardized units that permit comparisons between instruments.
Standardized Response Mean	Differs from effect size only in that the denominator is the standard deviation of change scores in the group (take account of variability in change scores rather than baseline scores)
Floor/ceiling effects	Floor (ceiling) effects occur when an instrument's lowest (highest) score is unable to assess a patient's level of ability, meaning that it is not possible to report most favorable or worst health states. Usually measured by recording the number and percentage of participants with the highest and lowest score.
Acceptability	How acceptable is the instrument in terms of completion by the participant – does it represent a burden? Can the assessment be completed by proxy? (Fitzpatrick et al. 1998)
Feasibility	Is the instrument easy to administer and process? (Consider extent of effort, burden, expense and disruption to staff/clinical care arising from administration of the instrument) (Fitzpatrick et al. 1998)

References:

Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technology Assessment 1998; 2(14): 1-73.

<http://www.ncbi.nlm.nih.gov/pubmed/9812244>

Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice, Third Edition. Pearson Prentice Hall, 2009.

Rehabilitation Measures Database. Statistics Review. Available at:

<http://www.rehabmeasures.org/rehabweb/rhstats.aspx>

Modified from <http://www.SCIREproject.com>

Outcome Measures: Glossary

Science Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research* 2002; 11: 193-205.

<http://www.ncbi.nlm.nih.gov/pubmed/12074258>

Salter K, Jutai JW, Teasell R, Foley NC, Bitensky J, Bayley M. Issues for selection of outcome measures in stroke rehabilitation: ICF activity. *Disability and Rehabilitation* 2005; 27(6): 315-340.

<http://www.ncbi.nlm.nih.gov/pubmed/16040533>

Stroke Engine Database. Glossary of Terms. Available at:

<http://strokengine.ca/assess/definitions-en.html>

Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, Bouter LM, de Vet HCW. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology* 2007; 60: 34-42.

<http://www.ncbi.nlm.nih.gov/pubmed/17161752>