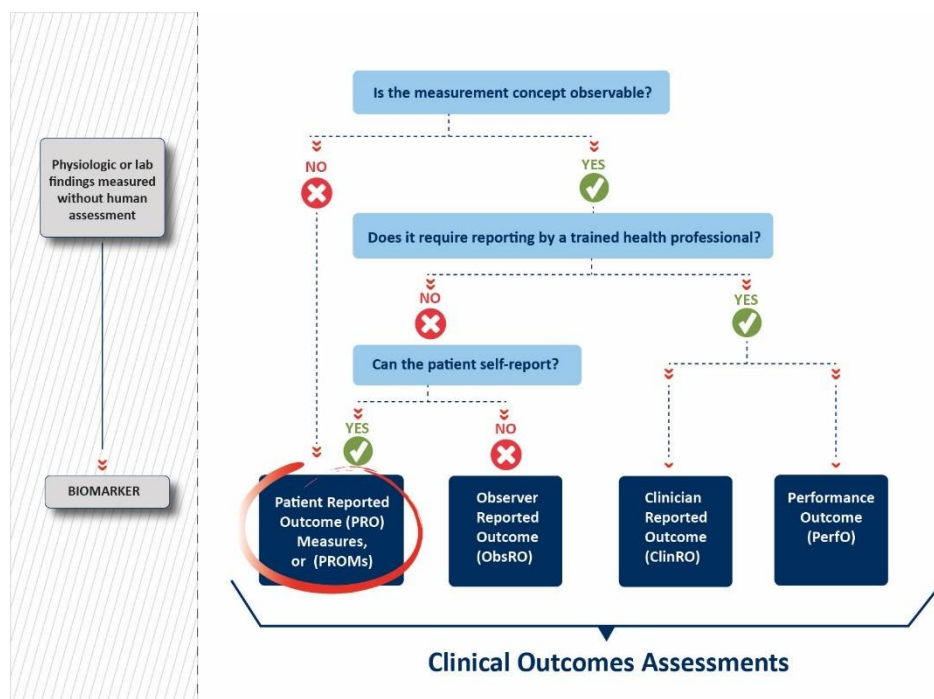


# Glossary of Health Status Measurement Terminology

**Ceiling and floor effects:** refer to the response range and the method of scoring an instrument. Thus an instrument, applied to a random sample of the population, which is not sensitive to lower levels of ill health and that is scored from 0 (good health) to 100 (poor health) would be said to manifest a floor effect, as most respondents would score zero. On the other hand if the instrument was scored from 0 (poor health) to 100 (good health) this would be referred to as a ceiling effect, as most respondents would score 100. Such floor and ceiling effects are more likely to be found in instruments with small numbers of items.

**Clinical trial:** an experiment to assess the efficacy of a treatment.

**COA:** The term Clinical Outcome Assessment or COA covers a number of health outcomes assessment instruments (questionnaires), including Patient Reported Outcome (PRO) measures (or PROMs, as used in the UK); Clinician Reported Outcome (ClinRO) measures; Performance Outcomes measures (PerfO's) and Observer-Reported Outcome (ObsRO) measures. Most of the health outcome measurement instruments that we manage and support are PRO measures.



**Construct:** a phenomenon that exists theoretically but that cannot be measured directly and is defined or operationalised in terms of other observed indicators. Thus it is not possible to measure depression in the same way as height, so indirect measures (questionnaires of self report, or clinical assessment of behaviour) have to be used.

**Construct validity:** where hypotheses are generated and a questionnaire tested to determine if it actually reflects these prior hypotheses. For example, the construct validity of the SF-36 has been checked to ensure that certain groups (e.g. older, lower social classes, those with illnesses) gain lower scores (i.e. indicating worse health) than other groups (e.g. younger, higher social classes, those without illnesses).

**Content validity:** the extent to which items on a questionnaire tap all the relevant aspects of the attribute they are intending to measure.

**Convergent and discriminant validity:** a measure should both converge with other indicators of the same concept and be able to discriminate unrelated indicators.

**Cost utility analysis:** a form of economic cost effectiveness analysis where the effects of health care interventions are assessed according to the quality adjusted life years gained or lost.

**Criterion validity:** the extent to which a measure correlates with a preexisting one, preferably a 'gold standard'. There are two types: (i) Concurrent validity: where a new measure is administered at the same time as a pre-existing one, and the two are correlated, and (ii) Predictive validity:

**Cronbach's Alpha:** A statistic used to determine the internal reliability of scales (see reliability).

**Dimensions of health:** theoretically or empirically distinct aspects of health, for example physical and mental health.

**Disease Specific Measures:** Questionnaires designed for use with a particular patient group.

**Domain:** see dimensions of health.

**Effect size:** a statistic for determining the difference between scores gained at two different times. This statistic has been recommended as a method of evaluating the sensitivity of health measurement instruments to important clinical change, calculated by dividing the mean change in score by the baseline standard.

**Face validity:** the need for a questionnaire to apparently tap, simply by item content, an underlying dimension. Questions should be unambiguous and easily understood, and should reflect issues appropriate to the dimension.

**Factor analysis:** A group of statistical techniques whose purpose is to produce a large number of variables to a smaller number of latent variables, i.e. variables that cannot be measured directly. Thus a number of questions may be seen as measuring the single theoretical concept of 'anxiety'.

**Floor effect:** see ceiling and floor effects.

**Generic measure:** a measure which is designed for use with any illness groups or population samples, as opposed to those intended for specific illness groups.

**Guttman Scaling:** A scale containing items, which are summed, that all tap aspects of the same phenomena. Each item contains sets of statements which are hierarchically ranked. Affirmation of a statement means that all statements below it are also affirmed and all those above it are not. Thus by affirming item number 3 in this example then 1 and 2 must also be true, whilst 4 is not: 4. I can walk very long distances beyond a mile; 3. I can walk a mile; 2. I can walk half a mile; 1. I can walk very short distances

**Health index:** where all the items of an instrument are summed producing one overall score.

**Health outcomes:** the end results of medical interventions and processes. These can be assessed in terms of mortality, morbidity, physiological measures and, increasingly, more subjective patient-based assessments of health.

**Health profile:** a questionnaire covering various dimensions of health, as opposed to a health index which sums all measured aspects of health into a single figure.

**Health related quality of life:** this refers to an individual's level of health related wellbeing. Measurement of health related quality of life addresses the various dimensions of health (see dimensions of health).

**Health status:** a level of health in terms of physical, social and mental wellbeing.

**Health:** there are numerous definitions of health, but perhaps the most widely quoted is that of the World Health Organisation which claims that health is a state of complete physical, mental, and social wellbeing and not merely the absence of disease or infirmity.

**Index of Responsiveness:** a statistic for determining the difference between scores gained at two different times.

**Instrument:** the tool with which health status or quality of life is measured, usually in the form of a questionnaire.

**Inter-rater reliability:** addresses the consistency of a measure when administered by different interviewers. This is tested by interviewing the same people with the same measures but using different interviewers with only a short period of time between. Kappa coefficient of agreement is the statistical tool used to assess whether differences were due to agreement or chance.

**Internal consistency reliability:** assessment of internal consistency reliability involves examining the extent to which a number of items addressing the same concept actually are doing so. There are a number of ways of calculating the correlation between items, for example: split-half reliability - whereby the measure is randomly split into two groups and reliability is assessed by the correlation between the two half tests, and Cronbach's Alpha - a statistical test of internal consistency based on the mean correlation between items.

**Internal consistency:** see internal consistency reliability.

**Interval scales:** It is assumed that data on an interval scale is ordered, and the distances between values on one part of the scale are equal in

distance to the distances between values on another part of the scale. Temperature is measured on such a scale. However interval scales lack an absolute baseline anchor point. For example, a thermometer is an interval scale but it is not possible to assume that 60°F is twice as hot as 30°F as 0°F is not as cold as it can get.

**Item content:** refers to the actual wording of the individual questions. Such content must at least satisfy requirements of face validity (see validity).

**Item:** an individual question which may stand alone or form part of a battery of questions in a dimension.

**Latent variable:** a variable that cannot be measured directly, but is measured indirectly (see also construct, factor analysis).

**Likert Scales:** A response scale in which respondents select from a range of options which are placed on a continuum, such as Never, Rarely, Sometimes, Often, Very Often. A number of related questions tapping the aspects of the same phenomenon will, ideally, have the same response categories and can be summed.

**Longitudinal study:** where individuals in a study are followed over time.

**Minimally important difference:** the minimally important difference is that difference in a score on health status measure that corresponds to the smallest change that patients consider important.

**Multi-dimensional measures:** instruments which consider health in more than one dimension/domain, of health; for example mobility, pain, mental health.

**Nominal scales:** these scales distinguish classes of objects. For example, the classification of sex into 1=male and 2=female is a nominal scale. A more complex nominal scale is the International Classification of Diseases, where numerical values classify all diagnoses and presenting problems. There is no hierarchy implied by the values ascribed, so in the example of sex it would be equally as valid to code 1=female and 2=male.

**Non-parametric methods:** statistical analyses which assume that data does not follow the normal distribution.

**Normal distribution:** Data that produces a 'bell curve' with the mean, mode and median all having the same value.

**Normative data:** data which are representative of a population.

**Ordinal scales:** Classes or objects are ordered on a continuum (for example, from Best to Worst). No indication is given as to the distance between values, although a hierarchy is assumed to exist. Thus when classifying an illness into 1=mild, 2=moderate and 3=severe it cannot be assumed that the extent of difference between mild and moderate is similar to the difference between moderate and severe.

**Parametric methods:** statistical analyses in which the data are assumed to be normally distributed.

**Precision:** the ability of an instrument to differentiate between illness groups, or states of health.

**Predictive validity:** the ability of an instrument to predict some other measure of outcome.

**QALY (quality-adjusted life year):** a generic measure of health benefit which attempts to represent the relative value attached by society to different improvements in health enabling systematic comparison between a variety of health care interventions. Comparisons between treatment programmes are expressed in quality-adjusted life years. With both a measure of the life years gained from a particular intervention and of the quality of life in each of those years it is possible to calculate the number of QALY's obtained. Thus an index of quality of life, multiplied by the number of years in that health state equals the number of QALY's.

**Random sample:** where each individual in the given population has an equal chance of selection into the sample.

**Ratio scales:** A ratio scale is an interval scale with an absolute zero point, so that ratios between values can be meaningfully defined; thus 10km is twice as long as 5km. Thus time, weight and height are all examples of ratio scales.

**Regression:** A group of statistical techniques whose purpose is to predict the value of some variable or variables when others are known.

**Reliability:** a reliable measure is one which produces consistent results from the same subjects at different times when there exists no evidence of change (see also test-retest reliability, internal consistency reliability, inter-rater reliability).



**Response range:** the set of answers available to respondents for each item.

**Responsiveness:** the extent to which an instrument can detect change in health status over time. (See also ceiling and floor effects, effect size, standardised response mean, index of responsiveness).

**Scales:** a graded system of categories. More frequently in psychology and health status measurement the term refers more specifically to a series of self report questions which can in some way be summed (see also nominal scales, ordinal scales, interval scales, ratio scales).

**Sensitivity to change:** an instrument's ability to detect change over time, sometimes referred to as 'responsiveness'.

**Standardised response mean:** a statistic for determining the difference between scores gained at two different times. It is calculated by dividing the mean change on a scale by the mean change in the standard deviation. Such a method is recommended when comparing the sensitivity to change of various health status measures (see also effect size).

**Subjective well-being:** the patient's assessment of their own health status as opposed to professionally or clinically defined indicators.

**Systematic review:** a review in which the methods for selecting and including or excluding publications are explicitly stated.

**Test-retest reliability:** this involves the administration of an instrument on two separate occasions to the same population. The correlation between scores provides an estimate of the measure's

reliability. The two occasions need to be far enough apart so that previous responses cannot be remembered but close enough in time so that change in the true score is minimal.

**Utility:** the preference for or desirability of a particular outcome in terms of health status.

**Validity:** the extent to which an instrument measures the desired underlying concept (see also face validity, content validity, criterion validity)

**Visual analogue scale:** a 10cm line on which the respondent indicates the intensity of his or her response. Phrases are printed at the ends of the line (e.g. 'no pain' and 'extreme pain') to indicate the scope of the scale.

**Weighting:** items which are given values indicating their relative importance to other items on a scale are said to be weighted. For example, on one measure, the Nottingham Health Profile, the item 'I cannot walk at all' is given a greater value than 'I need help to walk about outside'. There is a considerable body of evidence that in many instances such weighting schemes make only marginal differences to scoring simply by addition of raw scores.