Table 6: Clinimetrics/Psychometrics

RELIABILITY

Overall definition of reliability: The repeatability and stability (or internal consistency of a questionnaire) i.e. the extent to which a measurement is consistent and free from error.

Internal consistency

Definition: Assesses the homogeneity of the scale items.

Notes: Internal consistency is about whether the items within a subscale are measuring the same underlying construct. (eg, a depression questionnaire, a fixed set of tests of physical function). It is not relevant to unidimensional outcomes (eg, blood pressure, weight) or instruments that measure a set of unrelated constructs (eg, APGAR score, BODE index).

Statistics: It is generally examined using split-half reliability or Cronbach's alpha statistics. Item-to-item and item-to-scale correlations are also commonly used methods.
Split-half or Cronbach's alpha statistics: Excellent: ≥ 0.80; Adequate: 0.70 – 0.79; Poor: <0.70.
Note: caution alpha-values in excess of 0.90 may indicate redundancy. Inter-item and item-to-scale correlation co-efficients: adequate levels: Inter-item: between 0.3 and 0.9; Item-to-scale: between 0.2 and 0.9.

Notes: The main reason for using inter-item correlation rather than cronbach's alpha is for scales with a small no. of items e.g. <10 as short scales can give misleadingly low cronbach..

Test-retest reproducibility

Definition: When the instrument measures the same thing more than once, how consistent are the results? reproducibility addresses the degree to which the score is free from error (both systematic and random). Test-retest, intra-rater and inter-rater reliability all focus on this aspect of reliability.

Notes: Systematic error (such as the learning effect on the 6-minute walk test) can be reduced by doing a practice test.

Statistics: Test-retest or inter-rater reliability are commonly evaluated using correlation statistics including intraclass correlation coefficient (ICC), kappa coefficients (weighted or unweighted) (excellent: ≥ 0.75; adequate: 0.41 – 0.74; poor: ≤ 0.40; N.B. a minimum test-retest reliability of 0.80 is required if the measure is to be used to evaluate the ongoing progress of an individual in a treatment situation). 95% confidence interval, limits of agreement and Bland and Altman plots, coefficient of variation or standard error of measurement (SEM) (decision on cut-off points for acceptability is based on clinically meaningful differences)

VALIDITY

Overall definition of validity: Does the instrument measure what it purports to measure?

Criterion validity

Definitions: The following are all forms of criterion validity

- Concurrent validity: the degree to which a test correlates with a 'gold standard' criterion test known to be a valid test of the attribute of interest.
- Convergent validity: the degree to which a test correlates with another test which measures the same attribute (important when there is no established gold standard test to compare to).
- Predictive validity: Degree to which an attribute can be predicted using the result of a predictor test (e.g. the ability of an exercise test to predict peak oxygen uptake) or predict prognosis
- Discriminate validity: the degree to which it differentiates between groups of individuals known to differ in the attribute of interest (e.g. varying degrees of disease severity)

Definitions: The following are forms of validity for dichotomous outcomes

- Sensitivity: probability of positive result when the attribute of interest is present
- Specificity: probability of negative result when the attribute of interest is absent
- Positive predictive value: probability that the attribute is present given that the result is positive
- Negative predictive value: probability that the attribute is absent given that the result is negative

Statistics:

- Concurrent/convergent validity: correlations coefficients (0 to 0.25 = little or no relationship; 0.25 to 0.50 = fair relationship; 0.50 to 0.75 = moderate to good relationship; 0.75 to 1.0 = good to excellent relationship;
- Predictive validity: standard error of estimate (acceptable if below a level deemed clinically important); significance (and variability) of difference between estimated and measured attribute (acceptable if $p>0.05$, $CV<10\%$)
- Discriminate validity: Significance (and variability) of difference between groups of individuals known to differ in attribute of interest (acceptable if $p>0.05$; $CV<10\%$)
- Sensitivity, specificity and predictive values: percentage/ ratio statistics; receiver operating characteristic (ROC) analysis (graphic representation of balance between sensitivity and specificity). These are often used to establish cut-off points for diagnosing a condition and therefore acceptability depends on consequences of false negatives versus false positives. The point at which the curve turns usually represents the best choice of cut-off. Area under curve analysis (AUC): excellent: ≥ 0.90; adequate: $0.70 - 0.89$; poor: ≤ 0.70.

Construct validity (questionnaires, fixed batteries of physical tasks)

Definition: When there is no "gold standard" or other test to provide a basis of comparison. Convergent and divergent validity are considered to be forms of construct validity

- Convergent validity: The extent to which the scores on the questionnaire relate to other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that were measured. (i.e. constructs that theoretically **should** be related to each other are observed to be related to each other)
- Divergent validity: The extent to which constructs that theoretically should **not** be related to each other are observed **not** to be related to each other.

Statistics: 0 to 0.25 = little or no relationship; 0.25 to 0.50 = fair relationship; 0.50 to 0.75 = moderate to good relationship; 0.75 to 1.0 = good to excellent relationship; factor analysis

Content validity **(questionnaires, fixed batteries of physical tasks)**

- Face validity: The assumption of validity of a test based on its appearance as a reasonable measure of the attribute of interest
- Content validity: the extent to which the domain of interest is comprehensively sampled by the items of the questionnaire
  Notes: Face and content validity are usually assessed by non-statistical means (i.e. consultation with experts in the field or expert judgement)

RESPONSIVENESS

Definition: The ability to detect change in response to an intervention known to change the attribute of interest. Short term <1month, medium term 1 to 6 months, long term >6months.

Statistics: Commonly evaluated through correlation with other change scores such as effect sizes, standardised response means (mean change score divided by standard deviation) or relative efficiency (< 0.5 = small; 0.5 – 0.8 = moderate; > 0.8 = large). Sensitivity and specificity of change scores and receiver operating characteristic (ROC) analysis (i.e. has a minimal clinical important difference been detected when a real change in attribute has occurred-see validity section above for interpretation)

N.B. evaluation should also include evidence of floor and ceiling effects

Definition: Indicate limits to the range of detectable change beyond which no further improvement or deterioration can be noted

Statistics: Percentage/ratio (excellent: no floor or ceiling effects; adequate: ≤ 20% of patients who attain either the minimum (floor) or maximum (ceiling) score; poor: > 20% of patients who attain either the minimum (floor) or maximum (ceiling) score)

ARE THERE DATA ON NORMAL RANGES?

Please state if there are normal ranges established, including country of origin and age of subjects

ALL RELEVANT REFERENCES (PLEASE FORMAT AS PER JOURNAL OF CYSTIC FIBROSIS)

Manuscripts should use the 'Vancouver' style for references. For journal references, all authors should be included when there are six or less (first three only when seven or more), followed by the title of article, name of journal abbreviated according to British Standard 4148: 1975 (or left in full), year, volume, and first and last pages. For example:

1. Tockman MS, Anthonisen MD, Wright EC et al. Airways obstruction and the risk of lung cancer. Ann Intern Med 1987; 106:512-518.

For book references, the author(s) should be followed by the chapter title (if appropriate), editor(s) (if applicable), book title, place of publication, publisher, year and page numbers. For example:

2. Colby VT, Carrington CB. Infiltrative lung disease. In: Thurlbeck WM, ed. Pathology of the Lung. New York: Thieme Medical Publishers, 1988.

NB: For some instruments, you may find that no studies that have investigated its clinimetric properties in people with CF. In this situation, search for clinimetric studies that have been performed in healthy people or other patients with respiratory disease. Guide readers to these studies by stating "Some clinimetric properties of this instrument have been examined in healthy people (cite references) and/or patients with other lung disease (cite references)", but do not attempt to summarise their results.

Feasibility

Definition: Extent of effort, burden, expense and disruption to staff/clinical care arising from the administration of the instrument

EASE OF PERFORMANCE (for patients)
EASE OF ADMINISTRATION (for investigators, including scoring)
TIME TO ADMINISTER TEST (Include duration of a single test and whether a practice test is required)
SPACE REQUIRED
COST (Estimate one-off costs, e.g. equipment) (Estimate perpetual costs, e.g. maintenance, consumables)
SAFETY (List any hazards or side effects for the investigator and for the participant)
CAN THEY BE USED REPEATEDLY OVER TIME? (List any limitation on the number of times the instrument can be administered, e.g. cumulative radiation dose)
TRAINING OF RESEARCH STAFF (List whatever qualification or special training, if any, is required to administer the test)
AVAILABILITY OF STANDARDISED PROCEDURES (List any guidelines or protocols that are available to guide administration of the test)
SCORING (List any equipment, staff or software that are required or beneficial for scoring)
ALL RELEVANT REFERENCES (Please format as per *Journal of Cystic Fibrosis*, which uses the 'Vancouver' style)

Clinical Relevance

WHAT AGES HAS IT BEEN USED IN?
ALL RELEVANT REFERENCES (Please format as per *Journal of Cystic Fibrosis*, which uses the 'Vancouver' style)