**Online supplementary information DS2 – Definitions of psychometric properties and their corresponding indexes**

The following definitions are provided according to the COSMIN taxonomy of measurement properties.43

Internal consistency reliability: A form of reliability which is a measure of the degree of interrelatedness among the items in a rating scale. Measured using Cronbach’s Alpha or point biserial correlation coefficient (only for dichotomous variables).

Construct validity: The degree to which the content of a rating scale is an adequate reflection of the construct to be measured. This form of validity is not expressed as a statistical measure. It is evaluated by appraising if all items in a scale are deemed relevant for the construct, purpose of assessment and target population.

Convergent validity: The degree of correlation of a rating scale with existing scales measuring constructs that are theoretically related to the target construct. Correlation indexes are used, and expressed as Pearson’s r (Pearson product-moment correlation coefficient).

Discriminant validity: The degree of correlation of a rating scale with other scales which measure constructs that are anticipated to be relatively unrelated to the target construct. Correlation indexes are used, and expressed as Pearson’s r (Pearson product-moment correlation coefficient).

Structural validity: This refers to the factor structure of a scale and is obtained by conducting Principal Component Analysis, Exploratory Factor Analysis or Confirmatory Factor Analysis. Factors are described in terms of percentage of the total variance explained by each factor and by individual item loadings on one or more factors (coefficients of item loadings can vary between -1 and +1).

Criterion validity: This relates to correlation with a measure considered as “gold standard” to assess a given construct. Various indexes can be used to evaluate criterion validity. These include sensitivity, specificity and Area Under the ROC curve (AUROC) as detailed below.

Sensitivity: The proportion of people with the target condition (as diagnosed with the reference standard) who have a positive test result. Expressed as a value between 0 and 1, with values closer to 1 indicating greater accuracy.

Specificity: The proportion of people without the target condition (as diagnosed with the reference standard) who have a negative test results. Expressed as a value between 0 and 1, with values closer to 1 indicating greater accuracy.

Area under the ROC curve (AUROC): A parameter of overall test accuracy. Its value can vary between 0.5 and 1, with a value of 0.5 indicating that a test does not differentiate at all between subjects with and without the target condition and a value of 1 indicating a perfect test which screens positive all subjects with the target condition. This parameter will only be extracted for studies which determine post-hoc the cut-off score which provides the highest values for sensitivity and specificity.