

The SGRQ revisited: a psychometric evaluation

Muirne C.S. Paap, Cees A. W. Glas & Job van der Palen

University of Twente & Medisch Spectrum Twente, the Netherlands

THE SGRQ: LOOKING BACK

- The St. George's Respiratory Questionnaire (SGRQ) was developed in the early 1990's and consists of 50 items, of which 10 are polytomous and 40 dichotomous
- The sumscore is calculated using weights for each answering category (in total 76 weights):

$$\text{score} = 100 \times \frac{\text{summed weights for all positive items in the questionnaire}}{\text{sum of weights for all items in the questionnaire}}$$
- The weights were derived "empirically"; a group of patients was asked to rate the degree of distress they *would* experience for the situation described in the item, on a 10 cm Visual Analogue Scale (Quirk & Jones, 1990; Quirk, Baveystock, Wilson & Jones, 1991). Subsequently, these data were used to calculate the "average" experienced distress associated with a certain item/answering category and these were then used as weights from that point onward.
- Jones, Quirk, Baveystock & Littlejohns (1992) state that Principal-component analysis had resulted in 3 components, which could be interpreted as "Symptoms", "Impact", and "Activity". However, the manual provides equations for calculating both the total sumscore and subscale scores.
- In 2007, Meguro, Barley, Spencer & Jones developed a COPD-specific version of the SGRQ. They performed a Rasch analysis, which assumes a unidimensional structure, but did not explicitly test the factorial structure (the item fit of symptom, activity and impact items were only examined in relation to other items pertaining to the same subscale). If the Rasch model holds, the total unweighted sumscore is a good estimator for the position on the underlying construct. In this study, the empirical weights were not discussed or scrutinized.

THE SGRQ: SOME UNRESOLVED ISSUES

- How can the underlying dimensionality best be described? Can both the total score and subscale scores be used?
- Is a weighted score to be preferred over an unweighted score?
- Should certain items be removed? If so, does a reduction of the number of items lead to lower measurement precision (/reliability)?
- What construct is being measured?

BEFORE:

Multidimensional

Unidimensional

Weighted item scores

Unweighted item scores

Complete version

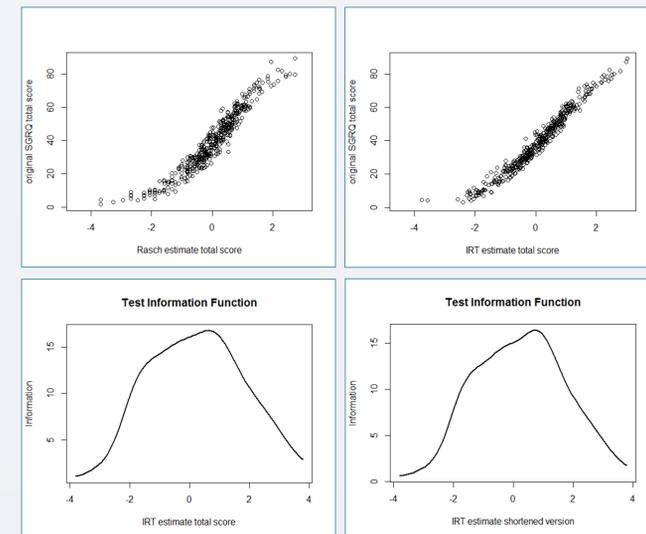
Shorter version (best items)

Distress

Subjective symptoms

METHODS & RESULTS

- Sample: 444 COPD patients (71% male) from two clinical trials conducted in the Netherlands; baseline data.
- Multidimensional Item Response Theory (MIRT) indicated that a MIRT model did not improve item fit compared to a unidimensional model. Mokken Scale Analysis (MSA) indicated that a unidimensional scale was adequate to describe the data.
- A unidimensional 2-parameter IRT model was compared to the Rasch model to assess whether weighted scores were to be preferred. The 2-PL showed better fit than the Rasch model.
- Several items showed poor fit. Test information did not change much after deletion of these items (see graphs).
- The 2PL estimates correlate highly with the SGRQ total score. This indicates that they measure the same construct, and distress is not adequately captured by the use of weights.



OUR FINDINGS INDICATE:

Multidimensional

Unidimensional

Weighted item scores (2PL)

Unweighted item scores

Complete version

Shorter version (best items)

Distress

Subjective symptoms