

Reliability of a longitudinal sequence of scale ratings

Annouschka Laenen¹, Ariel Alonso¹, Geert Molenberghs, and Tony Vangeneugden²

¹ Center for Statistics, Hasselt University, Belgium

² Tibotec, Johnson & Johnson, Mechelen, Belgium

Rating scales play an important role in many scientific areas and are mainly used when the trait of interest cannot be observed directly, such as the measurement of depression, anxiety, and quality of life. These instruments are often used within a longitudinal framework in view of studying patients' time course. One of the most important characteristics of such a scale is its reliability. This concept is related to the amount of measurement error, but can also be defined as the capacity of a scale to distinguish between subjects.

The field of psychometrics knows a long tradition in the study of reliability of, mainly, psychological and educational tests. In the 'classical test theory' reliability is defined as the true score variance divided by the total variance. However, the application of this expression is limited to a very narrow modeling framework. Generalizability theory was then developed to lift the reliability research out of these limitations and applied analysis of variance models with random effects to identify different sources of error. However, applying these models in a longitudinal framework often demands unrealistic assumptions, regarding the variance structure, error correlations and missing data pattern among others.

Linear mixed models, on the other hand, provide a framework that handles longitudinal data in a very flexible way. Laenen, Alonso, and Molenberghs (2007) extended the concept of reliability to this general modeling environment by defining four properties that every measure for reliability should fulfill. These authors identified a family of measures that fulfill the pre-specified set of properties, and therefore a family of reliability measures (Laenen *et al* 2008). Further exploration of some special members of this family revealed different interpretations for different members. Two measures appear to be especially useful: R_T , that expresses the average reliability over a sequence of repeated measurements, and R_Λ , that gives the reliability of the information we obtain by considering the whole sequence of measurements jointly. The latter measure shows how reliable information can be obtained even from scales that give rise to a considerable amount of measurement error, whenever repeated measures are considered. The methodology is applied to estimate the reliability of two different rating scales that are used to measure the severity of the symptoms in schizophrenic patients.

Key words: Reliability, Linear mixed models, Longitudinal data, Psychiatric rating scales.

Laenen, A., Alonso, A., and Molenberghs, G. (2007). A measure for the reliability of a rating scale based on longitudinal clinical trial data. *Psychometrika*, **73**, 443–448.

Laenen, A., Alonso, A., Molenberghs, G., and Vangeneugden, T. (2008). A family of parameters to investigate the reliability of a psychiatric symptom scale. *Submitted for publication*.