

## Editorial

### **Handle with care! Complex scoring systems in measuring health-related quality of life. Commentary on Wilson et al. "The SF-36 summary scales: problems and solutions"**

Measuring subjective phenomena such as perceived health status and health-related quality of life is not a trivial enterprise. In the absence of any gold standard, psychometric quality of methods for measuring these constructs can only be established using indirect and sometimes incomplete evidence from careful validation studies. Yet, since the days of Thurstone's proclamation that "attitudes can be measured"<sup>1</sup>, that even subjective phenomena are amenable to rigorous quantitative assessment, considerable progress was made in the development of instruments and techniques for controlling the uncertainty associated with subjective measurement. Modern multivariate scaling methods like factor analysis or item response theory models not only consolidate our understanding of what and how we are measuring, they also provide us with instructions on how to aggregate data from several items or parts of an instrument to obtain a meaningful composite score.

However, while these methods can substantially reduce our uncertainty concerning the structure and process of subjective measurement their complexity may at the same time introduce another type of uncertainty. Sophisticated and complex scaling methods tend to be less transparent for the researcher and may produce subtle but relevant problems which are difficult to detect. Careful inspection and evaluation of results are therefore needed when advanced scaling methods are applied.

In the current issue of *Social and Preventive Medicine*, Wilson and his colleagues<sup>2</sup> give an excellent example of how such a careful analysis provides valuable insights in the potential difficulties that may arise from uncritical adoption of complex scaling algorithms. Based on large data sets from population studies in Australia, they demonstrate that the usual method for calculating summary scores for the physical (PCS) and mental component (MCS) of the Short Form-36 Health Survey (SF-36)<sup>3</sup> produces scores not adequately reflecting the information contained in the subscales from which the summary scores are derived. Because physical subscales enter the MCS with negative weighting, a negative change in physical subscales may result in improved MCS scores even if no change in the mental subscales had occurred. The same problem is observed with the PCS to which the mental subscales make negative scoring contributions. Obviously, these inconsistencies may produce misleading results in both cross-sectional studies and studies assessing change over time. An alternative approach to the scoring of the SF-36 summary scales based on "structural equation modeling" is offered by the authors. It is shown that this approach avoids the problems of the traditional method while preserving the basic conceptual model of a physical and mental domain of health-related quality of life.

The paper by Wilson and his colleagues is especially noteworthy because the SF-36 is by far the most frequently used generic measure of health-related quality of life and has become widely adopted as a tool for evaluating and monitoring health outcomes in clinical and population-based studies. The SF-36 has been extensively (and successfully) tested by its developers from a variety of psychometric perspectives. Even though, problems of external validity and unexpected results as identified

and described by Wilson and his colleagues indicate that specific aspects of the traditional method of scoring the PCS and MCS still deserve further methodological evaluation.

It may be speculated that one reason for the rather late discovery of the problems was the relative complexity of the scaling model underlying the SF-36 summary scores. From a psychometric point of view, calculation of factor score coefficients from exploratory principal components analysis is a straightforward and in many situations well justified scoring method. In this case, however, some of the specific technical assumptions made by the developers (the presence of uncorrelated components, for example) appear to have resulted in by and large unexpected inconsistencies.

Given the current state-of-the-art in scoring and psychometric analysis of subjective measures and the enthusiasm arising from complex but promising developments in this area it is not easy to follow the advice given previously by Sir David Cox and his co-authors that statistical methods used in assessment of quality of life should be kept simple<sup>4</sup>. Interestingly, the structural equation modeling approach suggested by Wilson and his colleagues seems to be in principle closer to a simple unweighted composite score of the physical and mental subscales, respectively, than the traditional PCS/MCS scoring. It is perhaps a suitable combination of simplicity and adequate sophistication that guarantees results of subjective measurement to be not only methodologically meaningful but also meaningful to patients and other relevant decision makers in public health.

Thomas Kohlmann

#### References

1. Thurstone LL. Attitudes can be measured. *Am J Sociol* 1928; 33: 529–54.
2. Wilson D, Parsons J, Tucker G. The SF-36 summary scales: problems and solutions. *Soz Präventivmed* 2000; 45: 239–46.
3. Ware JE, Kosinski M, Keller SD. SF-36 physical and mental health summary scales: a user's manual. Boston, MA: The Health Institute, 1994.
4. Cox DR, Fitzpatrick R, Fletcher AE, Gore SM, Spiegelhalter DJ, Jones DR. Quality-of-life assessment: can we keep it simple? *J R Stat Soc Ser. A* 1992; 155: 353–93.