Modeling Subjective Health Outcomes

Top 10 Reasons to Use Thurstone's Method

Alberto Maydeu-Olivares, PhD, and Ulf Böckenholt, PhD†*

There is an increasing interest in measuring subjective health outcomes and in this issue of *Medical Care*, Krabbe[1] provides us with a much needed discussion of the potential usefulness of Thurstone's approach for quantifying subjective health outcomes using ordinal information, such as rankings. As pointed out eloquently by Krabbe, Thurstone's articles,[2,3] have had a tremendous influence on the development of methods for collecting and analyzing evaluative judgments in the form of preferences, attitudes, or more specifically, health outcomes. In fact, it is our view that Krabbe's discussion is rather modest. We feel more exuberant and present a list of top-10 reasons for using Thurstonian models. Although presented separately, these reasons should be viewed in conjunction because in this way they demonstrate the power and generality of this approach for studying and understanding evaluative judgments.

For many years, it was not possible to apply Thurstone's model in its full generality because of computational limitations. In technical terms, given rankings of p health outcomes p-1 dimensional integration over a normal distribution is needed if the model is to be estimated by maximum likelihood. Before the computer revolution, this integration was unfeasible and when Thurstone proposed his model more than 80 years ago only univariate integration could be done. Having been trained as an engineer, Thurstone[2] proposed an effective 2-stage estimation procedure that has been the skeleton upon which most classic research on Thurstonian models have been built.[4] Krabbe discusses this method very nicely. Given the proportions by which one health outcome is preferred over another (the P matrix in Krabbe's article), one obtains the corresponding normal deviates (Z values) by univariate integration. Then, the parameters of the model are obtained from the Z values by minimum distance methods. If only univariate integration is to be performed, only highly restrictive versions of Thurstone's general model can be estimated (such as the so-called Case V model considered by Krabbe).

Fortunately, these computational limitations are now overcome and Thurstone's model can be estimated in its full generality. As a result, the applicability of this approach has widened tremendously as our list of top-10 reasons illustrates. To put it bluntly, we are comparing a 1930s model car to the latest model when we look at Thurstone's classic Case V representation with its unrestricted counterpart. Still, the basic advantages of Thurstone's models should not be dismissed as illustrated by Krabbe, and our list of reasons highlights a number of them before we discuss the current capabilities of a Thurstonian analysis.

Reason 1. *It is easy for respondents*. Asking individuals to rate all outcomes under investigation on a sufficiently fine scale is cognitively a complex task and casts much doubt as to the reliability of such ratings. On the other hand, comparing outcomes requires less cognitive effort on the side of the respondents. One of the simplest approaches to gather comparative information is to solicit rankings of the health outcomes. In this case, health outcomes are compared directly with each other and not to the endpoints of a rating scale, which is both more meaningful and less sensitive to possible contextual effects induced by the use of arbitrary labels on a rating scale.

From the *Faculty of Psychology, University of Barcelona, Barcelona, Spain; and †Faculty of Management, McGill University, Montreal, Canada.
Reprints: Alberto Maydeu-Olivares, PhD, Faculty of Psychology, University of Barcelona, P. Valle de Hebrón, 171, 08035 Barcelona, Spain. E-mail: amaydeu@ub.edu.

ISSN: 0025-7079/08/4604-0346

Reason 2. *Inconsistencies can be modeled*. Both the use of ratings and rankings assume that respondents can order the health outcomes. This need not be so. Rather, because the evaluation of health outcomes is a complex task, it is prone to inconsistent judgments. Inconsistent judgments may be investigated using partial ranking techniques, such as the method of paired comparisons. Thurstone's model can be a useful diagnostic tool in both identifying and representing inconsistent judgments.[5] Specifically, it allows diagnosing health outcomes which are difficult to compare and identifying respondents who are systematically inconsistent and may have difficulties in evaluating health outcomes. To the best of our knowledge, this important individual-difference variable has received no attention in research involving medical applications.

Reason 3. *Both health outcomes and individuals may be scaled*. In addition to scaling health outcomes as illustrated by Krabbe (ie, estimating a metric scale for the outcomes), Thurstonian methods allow the scaling of individual respondents as well. Thus, although respondents need to provide only ordinal information, it is possible to estimate their underlying continuous responses. This approach seems more effective than asking respondents directly to provide a continuous rating which—because of its artificial nature—can be a challenging task. The estimated individual judgments can be used effectively in identifying the determinants of the subjective evaluations.

Reason 4. *It is a realistic and rich modeling framework*. In economics terms, Thurstone's model is a random utility model that assumes: (1) a utility maximization process within individuals, and (2) that the distribution of the utilities across individuals is normal. The parameters of the general Thurstonian model are the means in the population of the individuals' preferences for each health outcome (or mean utilities in economics terms), the variances of these preferences for each health outcome, and the covariances among preferences for different health outcomes. In applications, all these parameters are of interest. Thus, the utilities' variances provide us with valuable information about the degree to which respondents differ in their evaluation of health outcomes, whereas the covariances provide us with valuable information about the perceived similarity between health outcomes.

Reason 5. *Underlying (hidden) determinants of the valuations may be discovered*. When the number of outcomes to be judged is large we often wish to find a small set of underlying dimensions that account for the individual differences in the evaluative judgments. The position of the health outcomes can be depicted graphically along these dimensions, which allows for an easy interpretation of the underlying similarity structure of the health outcomes.[6]

Reason 6. *It is easy to do*. Thurstonian models for rankings of health outcomes can be easily estimated using standard structural equations modeling (SEM) programs with capabilities to perform mean and covariance structure analysis from ordinal data. The software Mplus,[7] for instance, has these capabilities, and Maydeu-Olivares and Böckenholt[8] provide code to estimate an array of Thurstonian models ranging from the unrestricted model to the most restrictive one (the so-called Case V model). Also, using SEM software greatly simplifies the scaling task. For instance, the mean utilities for the "death" and "perfect health" health states can be directly fixed to 0 and 1 to set the scale for the utilities of all other states. No reparameterization or normalization is needed.

Reason 7. *The validity of the inferences can be tested*. SEM programs provide goodness-of-fit tests[9] that provide information about the validity of the assumptions used and about the inferences drawn on the measurement of health outcomes. Thus, for example, a researcher can test easily whether the distributional assumptions are consistent with the data or whether judgments are reliable in their repeated evaluation of the items.

Reason 8. *Prior information about the health outcomes and/or the respondents can be incorporated into the model*. For instance, if a EQ-5D classification[10] has been used to derive the health states, this structure can be incorporated into a SEM model to determine the effect of each attribute level on the mean utilities.

Reason 9. *Nested and crossed sampling structures can be incorporated*. Thurstonian models may be viewed as 2-level models[11] (responses are nested within respondents). Further hierarchies may be included (individuals may be nested within countries) which would lead to a 3-level model. Alternatively, time may be included as a third level in the hierarchy.

Reason 10. *Numerous research opportunities!* By adopting a SEM approach, covariates may be included to explain individual differences in the valuations of the health outcomes, attitudes may be modeled simultaneously with the valuations, etc. An array of modeling opportunities and hence new substantive questions arise, many of which have not been explored up to now. When only valuations are being modeled, Maydeu-Olivares and Böckenholt provide a flow chart to guide researchers in selecting the appropriate model within the Thurstonian class. Thus, one should start investigating whether an unrestricted model provides a good fit to the data. If this is not the case, then probably the implicit assumption of a population homogeneous in its preferences for health outcomes is violated, and separate models for different subpopulations are needed. If, on the other hand, an unrestricted model provides a good fit, then it is worth investigating whether a model with independent utilities fits the data, as it is more parsimonious. We conjecture that in modeling health states this independence assumption is likely to be violated. If so, it may be possible to fit a dimensional model to the covariance structure (ie, a factor model) as these models facilitate the interpretation of the individual preferences' structure. Finally, if a satisfactory fit of a Thurstonian factor model is obtained, then a model where the mean utility values for the health states depend on the factors should be investigated. A satisfactory fit of this model provides an attractive interpretation about the source of the mean utilities.

In summary, Thurstonian models offer researchers an extremely powerful framework within which to not only measure—but also understand the source of—subjective health outcomes. Interest in the Thurstonian class of models is thriving[12–18] and we believe that by utilizing the full analytical power of these models the best is yet to come.

REFERENCES

1. Krabbe PFM. Thurstone scaling as a measurement method to quantify subjective health outcomes. *Med Care*. 2008;46:357–365.
2. Thurstone LL. A law of comparative judgment. *Psychol Rev*. 1927;79: 281–299.
3. Thurstone LL. Rank order as a psychological method. *J Exp Psychol*. 1931;14:187–201.
4. Bock RD, Jones LV. *The Measurement and Prediction of Judgment and Choice.* San Francisco, CA: Holden-Day; 1968.
5. Böckenholt U. Thresholds and intransitivies in pairwise judgments: a multilevel analysis. *Journal of Educational and Behavioral Statistics*. 2001;26:269–282.
6. Böckenholt U. Visualizing individual differences in pairwise preference data. *Food Quality and Preference*. 2006;17:179–187.
7. Muthén L, Muthén B. *Mplus.* Los Angeles, CA: Muthén & Muthén; 2006.
8. Maydeu-Olivares A, Böckenholt U. Structural equation modeling of paired comparisons and ranking data. *Psychol Methods*. 2005;10:285–304.
9. Maydeu-Olivares A. Limited information estimation and testing of discretized multivariate normal structural models. *Psychometrika*. 2006; 71:57–77.
10. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med.* 2001;33:337–343.
11. Böckenholt U. Hierarchical modeling of paired comparison data. *Psychol Methods*. 2001;6:49–66.
12. Maydeu-Olivares A. Thurstonian modeling of ranking data via mean and covariance structure analysis. *Psychometrika*. 1999;64:325–340.
13. Maydeu-Olivares A. Limited information estimation and testing of Thurstonian models for paired comparison data under multiple judgment sampling. *Psychometrika*. 2001;66:209–228.
14. Maydeu-Olivares A, Hernández A. Identification and small sample estimation of Thurstone's unrestricted model for paired comparisons data. *Multivariate Behavioral Research*. 2007;42:323–347.
15. Böckenholt U. Multivariate Thurstonian models. *Psychometrika*. 1990; 55:391–403.
16. Böckenholt U. Probability models and statistical analyses for ranking data. In: Fligner MA, Verducci JS, eds. *Applications of Thurstonian Models to Ranking Data*. New York, NY: Springer-Verlag; 1993.
17. Maydeu-Olivares A. Limited information estimation and testing of Thurstonian models for preference data. *Mathematical Social Sciences*. 2002;43:467–483.
18. Böckenholt U. Thurstonian-based analyses: past, present, and future utilities. *Psychometrika*. 2006;71:615–629.