Reviewer's report

Title: Individual response patterns determine time trade-off valuations of hypothetical health states: a secondary data analysis

Version: 1 Date: 16 December 2006

Reviewer: Ken Stein

Reviewer's report:

General

The subject of the paper is original and important. Little is known about the determinants of preferences, particularly on hypothetical states, and the authors are, in my opinion, correct to highlight the relative size of unexplained versus unexplained variance. The use of an internet enabled preference elicitation tool is also interesting and, while not entirely novel, is a potentially expanding area.

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

Was there any attempt to address potential effects arising from ordering of methods (i.e. VAS before TTO or vice versa). If the VAS was consistently presented before the choice based methods, the potential for anchoring heuristics might be considered. As the authors point out, the range in preference scores is remarkable. In some cases almost unbelievable e.g. equal lower limits for states as different as 112121 and 222223 or 332221 at 0.03. Are the authors sure that all respondents understood the task?

How was the list of “serious” medical conditions decided? The list is very restricted and some of the conditions are not necessarily that serious. An alternative might have been to ask for any long term limiting illness, in the manner that is employed in some (e.g. UK) national surveys.

Were there any effects discernible according to where/how the respondents completed the survey? In particular it would be interesting to know whether there were differences according to individual vs supervised completion.

Partly in relation to the previous point, it would be interesting to report whether there were differences in logical consistency according to setting. This basic test of the validity of the questionnaire/setting would provide reassurance and would be easily accomplished as there are several logically and unambiguously ordered pairs of health states within the sets valued.

The results of the TTO-LE vs TTO-10 comparison need further clarification. It is stated that the TTO-LE resulted in generally lower values than the TTO-10 and that this indicates a decreasing marginal utility of lifetime. It is not well argued why this is the case, though I suspect it relates to the fact that the sample all had greater LE than 10 years.

The G-study and cluster analyses are interesting. As the authors point out, little research has been carried out into the relative contribution of respondent and health state factors into elicited preferences. Previous research cited in this study showed much higher proportions of variance explained by health states. The authors should discuss potential reasons for these differences.

Insufficient details are given on the methods used in the cluster analysis. What algorithm was used? Was cluster analysis carried out on the principle components (presumably)? How was the number of clusters decided? Were the clusters stable, as the number of clusters was increased (i.e. does increasingly partitioning the data space retain most of the clusters as the n of clusters increase or does a completely different pattern emerge with each different n of clusters?).

The findings of the cluster analysis are intriguing. More explanation and discussion of the clusters is warranted as there are a lot and I am not certain that they all make sense. Clusters 1 and 2 clearly define people who are (for the TTO, at least) low raters and high raters. Otherwise, I am not very sure what the other 7 clusters tell me. This brings me back to my point about the number of clusters and how this was decided.
Finally, it might be helpful to order the health states according to severity, either on logical grounds or based on the overall mean preference scores (e.g. the 332221 state seems to be important in defining differences between clusters across methods).

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Discretionary Revisions (which the author can choose to ignore)

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:
'I declare that I have no competing interests'