Introduction and Objectives

The measurement of health related quality of life (HRQoL) from a patient perspective provides an important input to Health Technology Assessments (HTAs) as well as health economic evaluations using the cost per quality-adjusted life year (QALY) metric.

The present project was designed in response to the evidence that different generic (non-disease specific) instruments used to measure HRQoL produce different values for "utility" and in fact measure different constructs. Prior comparison studies suggest low levels of convergent validity - see Tables below:

This study is the first comprehensive comparison of the major HRQoL instruments based on multi-attribute utility (MAU) theory in Germany. It seeks to assess validity and content of existing generic MAU instruments and to identify the relationship between them.

Data and Methods

Empirical analyses make use of the German arm of the largest multi-instrument comparative survey done to date (MIC study) and analyzes a total of 1,269 participants from Germany.

Five generic instruments measuring HRQoL (15-D, AQoL, EQ-5D-5L, HUI-3, SF-6D) are considered alongside measures of well-being and capabilities.

MIC study respondents:

- Healthy population (263): representative in terms of age, gender, education.
-Patient population (1,009): asthma (147), cancer (115), depression (160), diabetes (140), heart problems (136), arthritis (159), chronic heart disease (152).

Results and Discussion

Pearson correlations between MAU Instruments:

<table>
<thead>
<tr>
<th></th>
<th>EQ-5D</th>
<th>HUI3</th>
<th>SF-6D</th>
<th>15D</th>
<th>AQoL-4D</th>
<th>AQoL-5D</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ-5D</td>
<td>1</td>
<td>0.79</td>
<td>0.60</td>
<td>0.79</td>
<td>0.79</td>
<td>0.79</td>
</tr>
<tr>
<td>HUI3</td>
<td>0.79</td>
<td>1</td>
<td>0.78</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>SF-6D</td>
<td>0.60</td>
<td>0.78</td>
<td>1</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
</tr>
<tr>
<td>15D</td>
<td>0.79</td>
<td>0.75</td>
<td>0.78</td>
<td>1</td>
<td>0.79</td>
<td>0.79</td>
</tr>
<tr>
<td>AQoL-4D</td>
<td>0.79</td>
<td>0.75</td>
<td>0.78</td>
<td>0.79</td>
<td>1</td>
<td>0.79</td>
</tr>
<tr>
<td>AQoL-5D</td>
<td>0.79</td>
<td>0.75</td>
<td>0.78</td>
<td>0.79</td>
<td>0.79</td>
<td>1</td>
</tr>
</tbody>
</table>

Correlation is significant at the 0.01 level (2-tailed).

Intraclass correlations with other MAU Instruments:

<table>
<thead>
<tr>
<th></th>
<th>EQ-5D</th>
<th>HUI3</th>
<th>SF-6D</th>
<th>15D</th>
<th>AQoL-4D</th>
<th>AQoL-5D</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ-5D</td>
<td>1</td>
<td>0.79</td>
<td>0.60</td>
<td>0.79</td>
<td>0.79</td>
<td>0.79</td>
</tr>
<tr>
<td>HUI3</td>
<td>0.79</td>
<td>1</td>
<td>0.78</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>SF-6D</td>
<td>0.60</td>
<td>0.78</td>
<td>1</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
</tr>
<tr>
<td>15D</td>
<td>0.79</td>
<td>0.75</td>
<td>0.78</td>
<td>1</td>
<td>0.79</td>
<td>0.79</td>
</tr>
<tr>
<td>AQoL-4D</td>
<td>0.79</td>
<td>0.75</td>
<td>0.78</td>
<td>0.79</td>
<td>1</td>
<td>0.79</td>
</tr>
<tr>
<td>AQoL-5D</td>
<td>0.79</td>
<td>0.75</td>
<td>0.78</td>
<td>0.79</td>
<td>0.79</td>
<td>1</td>
</tr>
</tbody>
</table>

Disaggregation by SF-36 Dimensions:

- **EQ-5D**
- **SF-6D**

Generic MAU instruments with choice-based scores:

- AQoL (Hawthorne et al. 1997; f.t.)
- EQ-5D (Dolan 1997; Shaw et al. 2005; f.t.)
- HUI-3 (Torrance 1988; Torrance et al. 1996; Feeny et al. 2002; f.t.)
- SF-12 / SF-6D (Brazier et al. 2002; Brazier and Roberts 2004; f.t.)

Generic Non-MAU instruments based on rating scale valuation:

- 15-D (Sintonen and Pekurin 1993; f.t.)
- QWB (Kaplan and Anderson 1988; f.t.)

Some early observations:

- Instruments are "not equivalent" (Mook and Kohnmann 2008).
- Instruments are "imprecisely related" (Fryback et al. 2010).

Edits healthy population:

- Answering time < 15 minutes.
- Visual Analogue Scale (VAS) < 70
- Confirmed a disease
- Duplicate inconsistency in EQ-5D mobility
- Inconsistency EQ-5D pain and AQoL-8D pain

Initially 344 healthy individuals, 260 total after edits

Edits patient population:

- Answering time < 15 minutes.
- Duplicate inconsistency in EQ-5D mobility
- Duplicate inconsistency in SF-36 overall health
- Inconsistency EQ-5D pain and AQoL-8D pain

Initially 1,216 patient individuals, 1,009 total after edits

Discussion:

Differences between MAU instruments in constructs, descriptive systems, and content sensitivity lead to differences in utility values. This suggests that, contrary to the impression generated by the generic term "utility", and by the use of their scores to calculate QALYs, the instruments are measuring different definitions of "health". Incremental "utilities" – forming the basis of conventional cost effectiveness analysis – may vary by up to 100 percent between MAU instruments, as shown by our linear geometric mean regression analyses.

This observation has potentially far-reaching implications – ranging from the choice of an appropriate (sensitive) instrument in clinical studies to limitations of the generalizability of health economic cost utility analyses.