Background

Cost-utility analyses need a measure to summarize the quality of life in a single index. The utility-preference approach of the EQ-5D-3L offers an interval measurement instrument resulting in an overall utility score while the psychometric approach of the SF-36 is based on a decomposed ordinal tool allowing to explore the various dimensions of the quality of life. Mapping technique can be used to obtain an utility score from the SF-36. No specific regression method has been recommended for implementing such mapping.

Objective

To compare the different regression methods for mapping the SF-36 into EQ-5D-3L based on French data.

Methods

EMOCAR Study design:
- French cohort of 904 patients with carotid endarterectomy followed from May 2011 to April 2016 with 3 visits (D0, D0+30, D0+120)
- EQ-5D-3L and SF-36 questionnaires collected for each patient at each visit

Steps:
1. Split the observations into two sets using random sampling : training and validation sets
2. Estimated regression in the training set
3. Implemented in the validation set to obtain utility score

Econometric methods:
- OLS: ordinary least square that relies on a quantitative variable to explain \( \rightarrow \) EQ-5D-3L index score
- Logit: multinomial logit model that estimates a qualitative variable \( \rightarrow \) each EQ-5D-3L dimension

Specification Two approaches concerning the explanatory variables: (1) summary score-based of the 8 dimensions of the SF-36 as quantitative variables; (2) all SF-36 items as dummy-independent categorical variables

Model performance (i) Predicted mean utility score; (ii) Mean absolute error (MAE) and mean squared errors (MSE); (iii) Distribution of errors

Results

<table>
<thead>
<tr>
<th>Parameters</th>
<th>OLS Specification 1 (score)</th>
<th>OLS Specification 2 (all items)</th>
<th>Logit Specification 1 (score)</th>
<th>Logit Specification 2 (all items)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Validation set (n=652)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.61</td>
<td>0.678</td>
<td>0.343-0.579</td>
<td>0.668-0.744</td>
</tr>
<tr>
<td>Mean (SD) EQ-5D-3L utility score</td>
<td>0.735 (0.185)</td>
<td>0.733 (0.20)</td>
<td>0.789 (0.193)</td>
<td>0.741 (0.241)</td>
</tr>
<tr>
<td>Min/Max EQ-5D-3L utility score</td>
<td>0.171/1.046</td>
<td>-0.076/1.075</td>
<td>-0.052/1</td>
<td>-0.199/1</td>
</tr>
<tr>
<td>MAE</td>
<td>0.113</td>
<td>0.116</td>
<td>0.116</td>
<td>0.126</td>
</tr>
<tr>
<td>MSE</td>
<td>0.024</td>
<td>0.026</td>
<td>0.032</td>
<td>0.038</td>
</tr>
<tr>
<td>P-value</td>
<td>0.9411</td>
<td>0.8193</td>
<td>&lt;0.0001</td>
<td>0.7074</td>
</tr>
</tbody>
</table>

Table 1: Performance of OLS and Logit models in predicting EQ-5D-3L scores

Distribution of errors
- The distribution of errors are quite similar between the four models
- OLS models estimated 66% of utility values with an absolute error > 0.05 but 44% with an absolute error > 0.1
- OLS model – specification 2 predicted almost 40% of utility values that are identical to the observed values
- Logit models estimated 60% of utility values with an absolute error > 0.05 and up to 47% with an absolute error > 0.1

Comparison by age
- OLS models predicted a better EQ-5D-3L score for the younger subgroups
- Logit models predicted a better EQ-5D-3L score for the older subgroups
- Specification 1 of the logit model predicted a significantly different EQ-5D-3L score for the older subgroups

Performance
- OLS models predicted a mean EQ-5D-3L utility score quite similar to the observed value
- Logit models predicted a mean EQ-5D-3L utility score higher than the observed value
- Logit models had a higher MSE value than OLS models
- Logit model with specification 1 predicted a EQ-5D-3L score significantly different from the EQ-5D-3L observed

Conclusion

Our study suggests that models using OLS method produce the best results for mapping SF-36 into EQ-5D-3L utility scores. Specification 2, using all items of SF-36, has also better performance. Logit model with specification 1 give poor conclusions.

Low EQ-5D-3L utility scores are poorly predicted with all models but this phenomenon is always observed in the existing literature.

References


Figure 1: Distribution of errors in EQ-5D-3L predicted scores

Figure 2: Observed and predicted EQ-5D-3L scores: comparison to OLS and Logit models

Observed vs. predicted
- OLS models predicted EQ-5D-3L score higher than 1 while logit models are limited to 1 (calculation of the utility score after regression with the French models
- OLS models predicted EQ-5D-3L scores closer to the observed score than logit models
- Low EQ-5D-3L scores are underestimated with all models