A comparison of direct and indirect methods for the estimation of health utilities from clinical outcomes

Mónica Hernández Alava, Allan Wailoo, Fred Wolfe and Kaleb Michaud
Introduction

• Mapping (or ‘cross-walking’) is used to estimate a utility score/index from a different outcome measure
  - clinical trials without a preference based measure
  - Within PROMS agenda as performance indicators
  - Essential element of VBP

• Mapping involves:
  • Estimating a relationship using a statistical model
  • Predicting using the estimated model

• **THIS IS ESSENTIALLY A STATISTICAL ISSUE!**
EQ-5D-3L UK-tariff

- Descriptive system 11111, 21123
  - 5 dimensions - mobility, self care, usual activities, pain, anxiety and depression
  - 3 levels in each dimension- no problems, some problems, extreme problems
  - 243 combinations
- Valuation (Dolan et al 1995) – utility scores
  - Analysis of preference data: 3000 individuals
Two general methods

• Direct: dependent variable – utility/index scores
  • 11213 -> 0.378

• Indirect: dependent variables – levels of descriptive system.
  • Expected index score is calculated as a second step
  • “Response mapping”
AIM: Estimate EQ-5D as a function of HAQ, Pain and other covariates – direct & indirect methods

- US not-for-profit database
- N=100,398 (16k patients)
- Adults with RA diagnosis
- Classic EQ-5D (UK tariff) distribution
- Multimodal
- Peak at 1
- Bounded top and bottom
- Gap between 1 and 0.883
Existing evidence

• Direct methods:
  • Linear regression
  • Tobit (often incorrectly applied!)
  • CLAD
  • Two-part models

Biased estimates of treatment effect
Methods and models

• Direct methods:
  • Adjusted Limited Dependent Variable Mixture Model
    (development of Hernández Alava et al 2012)
  • RE linear regression

• Indirect method:
  • Set of Generalised Ordered Probits
    (development of “Response Mapping” Gray et al 2006)
Direct method: Finite Mixture Modelling

- Useful where simple models don’t fit complex data
- Model data as a finite mixture of component models (usually of the same type)
- Often used where interest is in identifying clusters of groups
- But here we are interested in approach because of flexibility
- Any continuous distribution can be approximated by a mixture of normals
Mixture model example - 1 component
• Don’t need to use normal distributions
• More appropriate bespoke distribution
• Each component reflects EQ-5D properties
• Overcomes need for a class of “1”s
• Combination of:
  a) Adjusted dist
  AND
  b) Mixture framework
More formally...

\[ y_{it} \mid c = \begin{cases} 1 & \text{if } y_{it}^* \mid c > 0.883 \\ \max \{y_{it}^* \mid c, -0.594\} & \text{otherwise} \end{cases} \]

\[ y_{it}^* \mid c = x_{it}^T \beta_c + \varepsilon_{it}^c \]

\[ \beta_{0c} = z_i^T \alpha_c + u_i \]

\[ P(C_{it} = c \mid w_{it}) = \frac{\exp (w_{it}^T \partial_c)}{\sum_{s=1}^S \exp (w_{it}^T \partial_s)} \]

Limited at top, with a gap, and bottom

Within component (HAQ, HAQ^2, Pain, Age, Age^2)

Random effect (gender)

Component membership (HAQ, Pain, Pain^2)
Indirect method: Random Effects Generalised Ordered Probit

• 3 point ordered discrete dependent variable for each of the five dimensions of EQ-5D

• (RE) Ordered Probit – implicit parallel regression assumption too restrictive

• Multinomial logit model BUT ignores ordinality of the dependent variable
Indirect method: Random Effects Generalised Ordered Probits

- $q_{it}^s$ discrete dependent variables for $s=\{\text{mobility, self care, usual activities, pain, anxiety and depression}\}$

\[
\begin{align*}
P(q_{it}^s = 1|x_{it}, u_i^s) &= 1 - \Phi(x_{it}\beta_1^s + u_i^s) \\
P(q_{it}^s = 2|x_{it}, u_i^s) &= \Phi(x_{it}\beta_1^s + u_i^s) - \Phi(x_{it}\beta_2^s + u_i^s) \\
P(q_{it}^s = 3|x_{it}, u_i^s) &= \Phi(x_{it}\beta_2^s + u_i^s)
\end{align*}
\]

- Expected value calculated mathematically – average of all 243 utility values weighted by their estimated probabilities
Model selection and comparisons

- Explanatory variables: HAQ, HAQ², pain, gender, age and age²
- BIC to choose number of mixture components – 4
- MAE & RMSE (insensitive but widely used)
- Monte Carlo simulation to generate data from models and compare to observed data
EQ-5D distribution in each class and overall
Distribution of probabilities in each class
Table 4: Comparison of Models 1, 2 and 3

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>RE linear regression</th>
<th>ALDVMM</th>
<th>% diff 2 vs 1</th>
<th>RE GOProbit</th>
<th>% diff 3 vs 1</th>
<th>% diff 2 vs 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAQ 0-1</td>
<td>54,086</td>
<td>MAE 0.0968</td>
<td>0.0854</td>
<td>11.77%</td>
<td>RE 0.0906</td>
<td>6.46%</td>
<td>5.68%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE 0.1292</td>
<td>0.1215</td>
<td>5.96%</td>
<td>RE 0.1250</td>
<td>3.22%</td>
<td>2.83%</td>
</tr>
<tr>
<td>HAQ 1-2</td>
<td>38,307</td>
<td>MAE 0.1571</td>
<td>0.1458</td>
<td>7.17%</td>
<td>RE 0.1515</td>
<td>3.53%</td>
<td>3.77%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE 0.2061</td>
<td>0.2025</td>
<td>1.75%</td>
<td>RE 0.2033</td>
<td>1.39%</td>
<td>0.37%</td>
</tr>
<tr>
<td>HAQ 2-3</td>
<td>8,005</td>
<td>MAE 0.2309</td>
<td>0.2052</td>
<td>11.11%</td>
<td>RE 0.2130</td>
<td>7.77%</td>
<td>3.63%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE 0.2626</td>
<td>0.2520</td>
<td>4.01%</td>
<td>RE 0.2543</td>
<td>3.16%</td>
<td>0.88%</td>
</tr>
<tr>
<td>Overall</td>
<td>100,398</td>
<td>MAE 0.1305</td>
<td>0.1180</td>
<td>9.56%</td>
<td>RE 0.1236</td>
<td>5.30%</td>
<td>4.50%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE 0.1752</td>
<td>0.1693</td>
<td>3.37%</td>
<td>RE 0.1713</td>
<td>2.24%</td>
<td>1.16%</td>
</tr>
</tbody>
</table>
Only 1% with HAQ>2.5
Values exceed 1!
Does it matter?

- Example CE model: Rituximab for MTX intolerant patients with RA (Sharma et al. 2009)

<table>
<thead>
<tr>
<th></th>
<th>Inc Costs</th>
<th>Inc QALYs</th>
<th>ICER</th>
<th>%diff base lin</th>
<th>%diff base response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear model</td>
<td>£ 1,952</td>
<td>0.166</td>
<td>£11,754</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response</td>
<td>£ 1,952</td>
<td>0.190</td>
<td>£10,315</td>
<td>12.2%</td>
<td></td>
</tr>
<tr>
<td>Mixture</td>
<td>£ 1,952</td>
<td>0.158</td>
<td>£12,372</td>
<td>-5.3%</td>
<td>-19.9%</td>
</tr>
</tbody>
</table>
### Does it matter?

- Tentative results only
- Assume less severe patients

<table>
<thead>
<tr>
<th>Model</th>
<th>Inc Costs</th>
<th>Inc QALYs</th>
<th>ICER</th>
<th>%diff</th>
<th>%diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear model</td>
<td>£ 2,223</td>
<td>0.058</td>
<td>£ 38,441</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response</td>
<td>£ 2,222</td>
<td>0.062</td>
<td>£ 35,903</td>
<td>6.6%</td>
<td></td>
</tr>
<tr>
<td>Mixture</td>
<td>£ 2,225</td>
<td>0.066</td>
<td>£ 33,535</td>
<td>-12.8%</td>
<td>-6.6%</td>
</tr>
</tbody>
</table>
Conclusion/Discussion

• Linear models are not appropriate for mapping
  • Response mapping and mixture model approaches substantially better in all regards
  • …and it matters!

• Generalized ordered probit can be used for response mapping
  • Respects ordered nature of data
Conclusion/Discussion

• Bespoke mixture model performs best overall in this example
• Further work
  • Develop response mapping (correlations, more flexible functional forms)
  • Compare methods in other datasets/simulation/outcomes
  • How will it work with EQ-5D-5L?
    - Depends how valuations are modelled
References


To Discover And Understand.