ISSUE PANEL 3:
MEASURING UTILITY FOR ECONOMIC MODELS
WITHIN CLINICAL TRIALS: CAN WE DO BETTER?

Moderator:
Sorrel Wolowacz, PhD; Head, Health Economics Europe, RTI Health Solutions, Manchester, UK

Panelists:
Andrew Briggs, DPhil, MSc; William R. Lindsay Professor of Health Economics, Department of Health Economics & Health Technology Assessment, Institute of Health & Wellbeing, University of Glasgow, Glasgow, UK
Jennifer Petrillo, PhD; Associate Director, Global Health Economics & Outcomes Research, Novartis Pharmaceuticals Corporation, East Hanover, NJ, USA
Lynda Doward, MRes; European Head, Patient-Reported Outcomes, RTI Health Solutions, Manchester, UK
MEASURING UTILITY FOR ECONOMIC MODELS WITHIN CLINICAL TRIALS: CAN WE DO BETTER?

**PANEL:** Sorrel Wolowacz, PhD, RTI-HS; Jennifer Petrillo, PhD, Novartis; Lynda Doward, MRes, RTI-HS; Andrew Briggs, DPhil, University of Glasgow

**ISSUE:** Since NICE expressed a preference for utility estimates measured in patients using the EQ-5D (NICE, 2008), it has become increasingly common for EQ-5D data to be collected alongside clinical trials. However, in practice, such data often provide poor utility estimates for economic models for a number of reasons. Trials designed to evaluate efficacy and safety may not be appropriate or optimal for collection of utility estimates for economic modeling. For example, there may be limited opportunity to collect data for key health states relevant to the economic analysis within the trial follow-up (e.g., cancer trials often provide little data for patients after disease progression). The trial population may not represent patients in routine clinical practice which may introduce bias. For example, older patients or those with comorbidities or abnormal organ function are often excluded, resulting in a younger and fitter population than in routine clinical practice. Assessments are often made at regular scheduled visits, which may not coincide with the time during which events of interest (e.g., fractures or disease flares) affect quality of life. It is also crucial to consider whether EQ-5D is appropriate for the condition of interest (in terms of validity and responsiveness), and to design appropriate analyses of the data. Commonly, the mean utility, or the mean change from baseline, are reported at a series of time points. These data are often useless for economic models which usually require utility estimates for health states.

**OVERVIEW:** The panel will discuss approaches to determining whether to measure utility in a trial (versus an alternative study type), challenges of balancing the requirements for reimbursement against those of regulatory authorities in a single trial, whether EQ-5D is an appropriate measure (or an alternative measure is justified), optimal timing of assessments, and specification of analyses to utilise the power of patient-level utility data.
Issues With Utility Data Collected in Phase 3 Trials

- EQ-5D data are commonly being collected alongside clinical trials
  - Primary purpose is to provide utility weights for economic models
  - In practice, data collected has often had severe limitations

- Issues with trial design
  - Limitations in patient follow-up often results in key health states not being captured
  - Trial population may not be representative of patients in routine clinical practice
  - Number of utility assessments is often very limited
  - Assessments are often not scheduled to coincide with important events

- Issues with analyses
  - Analyses performed are often inappropriate for economic modelling
  - e.g., mean utility (or mean change from baseline) is presented at each scheduled assessment
  - i.e., analyses do not provide utility estimates for model health-states

- Issues with the utility instrument
  - Insufficient consideration of whether EQ-5D is appropriate in terms of validity and responsiveness in the indication being studied
Issues With Utility Data Collected in Phase 3 Trials

- Panel discussion
  - Approaches to determining whether to measure utility in a planned trial (versus performing a separate study)
  - Optimising the design of the utility data collection in the trial, e.g., selection of the utility instrument, patient follow-up, number and timing of assessments
  - Challenges of balancing the requirements for reimbursement against those of regulatory authorities in a single trial
  - Specification of analyses to utilise the power of patient-level utility data in economic models

- Audience participation / discussion session
# Panel Perspectives

<table>
<thead>
<tr>
<th>Name</th>
<th>Perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jennifer Petrillo</td>
<td>Industry perspective - health economics &amp; outcomes research</td>
</tr>
<tr>
<td>Lynda Doward</td>
<td>Patient-reported outcomes research</td>
</tr>
<tr>
<td>Andy Briggs</td>
<td>Health economic modelling</td>
</tr>
<tr>
<td>RTI-HS</td>
<td></td>
</tr>
<tr>
<td>Glasgow University</td>
<td></td>
</tr>
</tbody>
</table>
Issues With Utility Data: Industry Perspective

Jennifer Petrillo, PhD
4 November 2013
Why Are Utilities So Important...

Step 1
Obtain utilities from trials (EQ-5D or other preference-based measure)

Step 2
Run cost-effectiveness model with trial-based utilities to obtain incremental cost-effectiveness ratio (ICER)

Step 3
Submit ICER evidence to reimbursement bodies for acceptance of product

Intervention administered

Patients complete EQ-5D to reflect utility

Average trial sample ~300 patients

Utilities input to cost-effectiveness model

ICER

(hopefully) NICE decision
NICE (P)reference Case: A Hierarchy of Approaches

EQ-5D Utilities

I: Generic utilities
- Mapping to EQ-5D

II: Condition-specific utilities
- HUI or SF-6D
- Direct valuation of CSM

III: Vignettes
- From CSM
- From Lit/KOL/Patients

CSPBM
How Are Utilities Currently Collected? Example 1

- Phase II dose ranging study in PD-LID
  - Assessment at Baseline and Week 9
  - Modified Abnormal Involuntary Movement Scale (mAIMS)
  - EQ-5D

Table 2. Change in mAIMS, EQ-5D utility, and EQ-5D VAS scores between Baseline and Week 9

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean (SD)</th>
<th>Min</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Δ mAIMS</td>
<td>-4.8 (5.2)</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>8</td>
<td>19</td>
</tr>
<tr>
<td>Δ EQ-5D utility</td>
<td>0.024 (0.255)</td>
<td>-0.820</td>
<td>-0.104</td>
<td>0.000</td>
<td>0.104</td>
<td>0.768</td>
</tr>
<tr>
<td>Δ EQ-5D VAS</td>
<td>-0.05 (16.07)</td>
<td>-45</td>
<td>-10</td>
<td>0</td>
<td>10</td>
<td>41</td>
</tr>
</tbody>
</table>

- **Issues**: Small sample, uncertain dose / treatment regimen, not powered for differences, PD-LID concepts may be unrelated to EQ-5D

While mAIMS scores improved by an average of 4.8 points, mean change in the EQ-5D utility score and the EQ-5D VAS were only 0.024 and -0.05, respectively (Table 2).

Presented at ISPOR 18th Annual International Meeting, May 18-22, 2013, New Orleans, LA.
How Are Utilities Currently Collected? Example 2

<table>
<thead>
<tr>
<th>Health state defined by BCVA category (letters; treated eye)</th>
<th>RESTORE*</th>
<th>Lloyd et al**</th>
<th>WSE RESTORE*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean utility (SE)</td>
<td>Mean utility (SE)</td>
<td>Mean utility (SE)</td>
</tr>
<tr>
<td>1: 80–100</td>
<td>0.860 (0.034)</td>
<td>0.830</td>
<td></td>
</tr>
<tr>
<td>2: 76–85</td>
<td>0.860 (0.014)</td>
<td>0.750</td>
<td>0.866 (0.024)**</td>
</tr>
<tr>
<td>3: 66–75</td>
<td>0.813 (0.012)</td>
<td>0.750</td>
<td>0.831 (0.025)</td>
</tr>
<tr>
<td>4: 56–65</td>
<td>0.802 (0.014)</td>
<td>0.715</td>
<td>0.817 (0.028)</td>
</tr>
<tr>
<td>5: 46–55</td>
<td>0.770 (0.018)</td>
<td>0.680</td>
<td>0.789 (0.035)</td>
</tr>
<tr>
<td>6: 30–45</td>
<td>0.700 (0.027)</td>
<td>0.680</td>
<td>0.728 (0.044)</td>
</tr>
<tr>
<td>7: 26–35</td>
<td>0.681 (0.053)</td>
<td>0.530</td>
<td>0.487 (0.080)</td>
</tr>
<tr>
<td>8: 0–25</td>
<td>0.547 (0.083)</td>
<td>0.340</td>
<td>0.785 (0.115)</td>
</tr>
</tbody>
</table>

*Health state index reported by patients in RESTORE using the EuroQol 5D (EQ-5D) questionnaire. Mean utility (index of health) was measured for each health state (defined by visual acuity in the treated eye). Means were calculated using a regression technique for repeated measurements at baseline, month 3, month 6, and month 12. It was necessary to pool data from several measurement points to be able to cover all possible health state transitions with a sufficient sample size. Possible trend effect in the pooled data was adjusted (p < 0.05).

**80–100 letters and 76–85 letters combined due to small sample size > 85 letters

BCVA, best corrected visual acuity; BSE, better seeing eye; SE, standard error; WSE, worse seeing eye.

Issues: Not enough patients at each of the model states; little to no utility difference in some states

Limitations Across Trials

**WHO**
- Trial population may not be representative of patients in routine clinical practice
- Severity ranges, comorbidities, acute events all could be missed

**WHAT**
- Treatment interventions geared towards symptom/biomarker reduction which may not allow improvement in functioning and QoL in a single trial
- Measurement concepts may not be compatible with the EQ-5D

**WHEN**
- Study design is focused on clinically meaningful time points
  - May be short- or longer-term focused
  - Components of the treatment response (and measurement timing) may not correlate with changes on preference-based measures
- Need time to allow impact on utilities (QoL)
Health-related quality of life (HRQoL) defined as “the patient’s subjective perception of the impact of his disease and its treatment(s) on his daily life, physical, psychological and social functioning and well-being” (EMA, 2005)

Cost-utility analysis seeks to establish whether differences in expected costs between treatments can be justified in terms of changes in expected health effects.

Health effects are expressed in terms of quality-adjusted life-years (QALYs)

- The product of survival (or time in a state) and the quality of that state (measured as a utility)
- Requires a (health-related) quality-of-life ‘utility’ (weight)
Taking the Patient Perspective...

• Utility data collected via clinical trials
• So what’s the problem?
  – Is the trial population representative?
  – Do we have the right instrument?
  – Does the timing of assessments make sense?
• If we are asking patients to provide data via clinical trials, we need to be sure that those data are worth collecting
  – Are we doing the best we can?
Are Trial Populations Representative?

- Patients enrolled onto clinical trials may not be representative of patient population

- Age
- Diagnostic subgroup
- Disease severity
- Comorbidities
- Geographic location
- Socio-economic status
- Race / ethnicity
- Language
Underrepresentation in Trials

- Only 2.5% of cancer patients are enrolled into NCI-funded clinical trials (Ford et al., 2007)
  - Groups underrepresented included
    - African American men, Latinos / Hispanics, Asian and Pacific Islanders, American Indians / Alaska natives
    - Adolescents and older adults (aged ≥ 65 years)
    - Residents of rural areas
    - Low socioeconomic status groups

- Patients enrolled in clinical trials for Motor Neurone Disease (amyotrophic lateral sclerosis) may not represent the patient population (Chio et al., 2011)
  - Trial participants are younger, have longer diagnostic delay, more likely to have spinal involvement, more likely to be male
Language

- Trial populations tend to include the official or majority languages
  - What about groups whose first language is a minority language?
    - Regional, indigenous languages (e.g., Welsh, Breton, Basque)
    - Languages from other areas (e.g., Turkish spoken in the UK, Estonian spoken in Sweden)
    - Minority language of sovereign state may be a majority language of a local regional area
  - What about groups whose first language is a community language?
    - Languages spoken by migrant communities
    - Largest number of community languages in Europe can be found in the UK: over 300 languages are currently spoken in London schools
Impact of Non-representative Population

• Lack of diversity in randomised study populations reduces opportunities for discovering health effects that may be particularly relevant to underrepresented populations

• In terms of utilities…
  – The ‘weight’ component of QALYs may be skewed!
  – The health effects expressed by the QALY represent those for a (usually) fitter, younger, less socially, economically, and racially diverse population
  – Cost-effectiveness arguments generated may be flawed!

• What are we missing?
Do We Have the Best Instrument?

• How do we judge what is ‘the best’?
  – Acceptability to HTA body?
    • EQ-5D / Health Utilities Index (HUI-2 and HUI-3) / SF-6D
  – Suitability for therapeutic area from clinical and patient perspective?
    • Does it cover relevant concepts?
    • Does it cover irrelevant concepts?
    • Is there anything missing?
  – Does the instrument recall period make sense in the context of the therapeutic area and trial design?
Are the Questionnaires We Use Relevant?

- Different methods used to measure HRQoL produce different utility values
  - Understandably, HTAs like consistency to allow comparison across appraisals
  - Preference for generic measures
- Content coverage more relevant for therapeutic areas where physical limitations are an important disease feature
  - e.g., EQ-5D includes domains on mobility, self-care, usual activities, pain / discomfort and anxiety / depression
    - How can we ensure that we capture key symptoms are not addressed by measure?
    - How relevant are these for therapeutic areas where social, relationship, and emotional issues are a key feature of disease?
Timing of Assessments

• Assessment of outcome in chronic episodic conditions is always a challenge!
  – Multiple sclerosis (relapsing-remitting)
    • Rate of flare-ups: 2.1 ± 1.2 per year (Fernández-Megía et al., 2010)
    • Clinical trial duration: 12-18 months
    • Number of assessments: 2
  – Chronic obstructive pulmonary disease
    • Rate of exacerbation 4.6 per year (physician-reported mean) (Kessler et al., 2006)
    • Clinical trial duration: 8-12 weeks
    • Number of assessments: 2

• How well are we capturing the episodes of interest?
Key Question

• How do we optimise collection of utility data in the context of a clinical trial to ensure that the utility values we use in economic modelling are the most realistic and representative?
References


Analysing utility data from clinical trials
Andrew Briggs
William R Lindsay Chair of Health Economics

Overview

- Traditional approach to analysing trial data
- Event based analysis as alternative
- UKPDS example
- EVOLVE example
Traditional approach

- Analyse utility data by clinical trial arm
- Direct utility or change from baseline
- Test differences in utility between arms

Event-based approach

- Analyse relationship between clinical events and utility
- Combine difference between events and utility given event to estimate utility difference
- Conditional independence
UKPDS example

- 5000+ patients followed for median 11 years
- Significant difference in long-term adverse events of diabetes with treatment
- Cross sectional survey of EQ5D (n=3667)
- No significant difference between arms

Derived from: Clarke et al, MDM, 2002
EVOLVE example

- 3547 of 3668 patients had EQ5D measured
- Longitudinal data
- Little significant difference between arms
- Highly significant GEE regression

Briggs et al 2013 presentation at ASN

Summary

- Analysing utilities by arm suffers from lack of power
- An event based analysis may be more powerful
- Allows establishment of direct and indirect effects of treatment on utility
- Pre-specification will help communication?
Summing up

- Consider whether the planned trial is appropriate for utility measurement
  - Is it feasible to observe key model health states in the trial?
  - Is direct measurement of utility in patients experiencing key model health states feasible?
  - Can a sufficient number of assessments be included?
  - Can assessments be timed to coincide with the events of interest?
  - Is the trial population representative of the population in routine clinical practice? Could excluded patients who would be eligible for treatment be followed up for utility?
  - Would an observational study (or a combination of the trial and an observational study) be more appropriate?

- Consider whether EQ-5D is the most appropriate instrument
  - Is EQ-5D valid and responsive in this indication?

- Consider the optimal design of utility assessments in the trial
  - Number and timing of assessments
  - Patient follow-up (e.g., after progression, excluded patients)

- Consider what analyses should be specified
  - Align with model health states / events
  - Optimise sensitivity, e.g., exploring association between change from baseline and continuous (rather than categorical) clinical variables
  - Capture correlation between better and worse health states (e.g., using regression modelling)
Discussion Session

- What issues have you encountered in using utility data collected in trials in economic models?
- Or maybe you disagree with the panel and feel everything is great – let us know!
- What recommendations would you make for improvements?
- What barriers to optimal design of data collection and analysis have you encountered?
- Would an ISPOR Good Practice publication be appropriate and helpful?