
Taskforce Representatives:
Suzy Paisley (Moderator)
Hélène Chevrou-Séverac
Andrew Lloyd
Roberta Ara
OBJECTIVE: To develop guideline recommendations for good practices when

1) identifying, reviewing and synthesising evidence from the literature &
2) using the HSU estimates in cost-effectiveness models in health care.
Polling and Q&A

ISPOR Congress Platform

Mobile App
Search “ISPOR” in the App Store or on Google Play!

WiFi Network: ISPORGlasgow | Password: IQVIA2017

https://myispor.cnf.io/

Step 1
Step 2
Step 3
(Select the appropriate session)

F7: HEALTH STATE UTILITY (HSU) IDENTIFICATION AND USE IN COST-EFFECTIVENESS IN DECISION MODELLING - AN ISPOR TASK FORCE
Poll: Do HSU estimates from RCTs constitute the best evidence for cost-effectiveness models?
Today’s session

• Experience from industry
• Search, review and appraise
• Synthesis of HSU evidence
Reporting Utilities used in Cost-Effectiveness Analysis: Experience from the industry

Challenges in reporting Health State Utility Values (HSUVs) – experiences from the industry

Dr. Hélène Chevrou-Séverac

HEOR Director, Medical Affairs, Celgene International

European ISPOR conference 2017, Glasgow
Challenges in reporting HSUVs – experiences from the industry

- Theory versus Practice of generating/using HSUVs for analytical decision models?
- How are HSUVs reported in models developed by manufacturers?
- Can we do better?
Challenges in reporting HSUVs – experiences from the industry

- Theory versus Practice of generating/using HSUVs for analytical decision models?
- How are HSUVs reported in models developed by manufacturers?
- Can we do better?
Theory of generating and reporting HSUVs

**Phase I**
- Human Pharmacology
  - ‘First in Human’

**Phase II**
- Therapeutic Exploratory:
  - Indication clearer
  - Therapeutic effect
  - Safety (toxicity)
  - Proof of concept

**Phase III**
- Confirmation of efficacy and safety in:
  - Specific indication
  - Randomized trial
  - Double blinded
  - With an active comparator

**Phase IV**
- Post-approval studies:
  - PASS
  - Real world data

---

Exploration of the disease and treatment pathway of the expected targeted disease / indication to understand the possible health states.

Exploration of CEA model structure, possible health states, HR-QoL & HSUVs published about the indication from SLR; existing registries/RWD studies.

-Trial design matching expectation of HTA agencies and inclusion of HTA-required outcomes (including HR-QoL) into the trial protocol; Parallel RWD study to capture better HSUV by health state.

Launch sequences in countries, with CUA model adaptation, including HSUV calculation based on local country-weight/algorithm.

Regulatory and HTA dossiers ready pre-launch, with CUA model.
Some practices in generating and reporting HSUVs

**Phase I**
- Human Pharmacology
- ‘First in Human’

**Phase II**
- Therapeutic Exploratory:
  - Indication clearer
  - Therapeutic effect
  - Safety (toxicity)
  - Proof of concept

**Phase III**
- Confirmation of efficacy and safety in:
  - Specific indication
  - Randomized trial
  - Double blinded
  - With an active comparator

**Phase IV**
- Post-approval studies:
  - PASS
  - Real world data

- Rarely: early CEA model development; no influence of the HEOR or PMA team on the GO/NO GO decision in Phase III trial
- Some exploration of the disease and treatment pathway - not always involvement of HEOR/PMA

- Occasionally: early CEA model development; no influence of the HEOR or PMA team on the GO/NO GO decision in Phase III trial

- No always HR QoL instruments included, or HSUV, not often in the longer term; not always knowledge about how to analyze them

- In a hurry:
  - Literature search on HSUVs
  - RW study
  - Vignette study
  - Mapping if disease-specific PRO

**Need to catch-up on all what is missing for the regulatory and HTA submission**
Missing HSUVs from het phase 3 randomized clinical trial program

- **Cases when it can happen:**
  - Start-ups with lack of awareness of HTA submissions or without HTA experts at the time of the development of the phase 3 clinical trial design
  - When the study didn’t include an active comparator required in some HTA submissions; and the RCT from the comparator didn’t collect HSUVs (a comparator still seen as SoC even after 5-10 years)
  - When it doesn’t make sense to collect HSUVs as part of the RCT

- **Solutions:**
  - The quickest way is a systematic literature search on all HSUVs for the targeted indication coming from any type of publication (RCTs or RWD)
  - To generate the missing data by different methods:
    - Mapping of a disease-specific PRO into an HSUV
    - RWD study to collect the HSUV in patients from the overall patients population or a patients sub group using the new drug (in research)
    - Do a vignette study to gather HSUV
Challenges in reporting HSUVs – experiences from the industry

- Theory versus Practice of generating/using HSUVs for analytical decision models?
- How are HSUVs reported in models developed by manufacturers?
- Can we do better?
Justifying model choice and inputted data is key in HTA submissions

“The rigour of the selection process in modelling lies not in adherence to pre-defined criteria but in: justifying, testing and making transparent the judgements underlying selection decisions.”

When reporting HSUVs from randomized clinical trials, it is expected that:
- First, the SAP includes right algorithm to calculate the HSVUs from the instrument collected
- And second, the SAP should consider analyses of the HSUVs changes from baseline, difference between arms, difference over time; as well as full statistical analysis by health states fitting the CEA model structure

Unfortunately the HSUVs are not always reported appropriately in the 1st publication of the results of the RCTs

Issues often encountered:
- Trial not powered to demonstrate a significant difference in HR-QoL between the arms
- HSUVs not collected by health state; instead collected in each arm at planned visits fitting the capture of the clinical primary/secondary endpoints => might miss patients in extreme/severe health states?
- Time of RCT planned visits for collecting events might not match the occurrence of disease progression (ex: flares in CD?) and changes in HSUVs?
- Lack of full statistical analysis (95% CI not always calculated); lack of information on the clinical MID
- Utilities parameters are as well rarely tested into Probability Sensitivity Analyses (PSA)
Any recommendations from HTA agencies for HSUVs?

The NICE Reference Case\(^1\) requires that evidence to inform parameters of clinical effectiveness should be identified by systematic review. In its specification or definition of ‘systematic review’ for these parameters, the NICE Methods Guide refers to the systematic review methods of the Centre for Review and Dissemination (CRD).\(^1\) For all other types of evidence, including utilities, costs, and baseline risks of events, the need to be systematic and transparent is specified several times but the requirement for a systematic review of these types of evidence is not specified. There is an implication that a systematic and transparent process should be used but that this should not or cannot necessarily adhere to conventional systematic review methods.
Feedback from DSU-TSD 12 NICE 2011: “The authors reported a wide range of methodological variation in the use of utility values and a lack of clarity in the reporting of detailed methods used in the submissions.”

Source: “NICE DSU TECHNICAL SUPPORT DOCUMENT 12: THE USE OF HEALTH STATE UTILITY VALUES IN DECISION MODELS”; 2011; Ara and Wailoo

Real-life check:
- “Health state utilities for remission, mild disease, and moderate-to-severe disease were obtained from a pair of studies by Tsai et al and Punekar and Hawkins,(10,11) which presented utility weights based on EuroQol five dimensions data from a UK population.” (Wilson et al., 2017)

<table>
<thead>
<tr>
<th>Health state costs and utility weights</th>
<th>Cost/utility by health state</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>Remission</td>
</tr>
<tr>
<td>Per-cycle cost</td>
<td>£240.49</td>
</tr>
<tr>
<td>Per-cycle utility weight</td>
<td>0.88</td>
</tr>
</tbody>
</table>

- “Treatment-specific adverse event rates, along with utility decrements for selected events (eg, serious infection, tuberculosis, lymphoma, hypersensitivity reactions, and skin reactions), were obtained from the published literature.”
Previous example: “Health state utilities for remission, mild disease, and moderate-to-severe disease were obtained from a pair of studies by Tsai et al and Punekar and Hawkins,(10,11) ….” – Wilson et al. (2017)

Publications referenced (10 and 11):
- Tsay et al. (2008): “The primary effectiveness measure … was the QALY. This benefit was quantified using published data on health state preferences obtained from a UC patient survey carried out in Cardiff Hospital using the EQ-5D [13, 17].”
- Punekar et al. (201): “The intermediate treatment outcomes of colectomy, symptom-free remission and surgical complications were translated into the final outcome of QALYs using the health state preferences obtained from a UC patient survey carried out in Cardiff Hospital using the EQ-5D [16] and valued using UK tariffs [17].”

Initial publications used by the 2 publications above:
- One is an abstract on CD, while the indication studies was UC; and the abstract doesn’t include any Utilities data—maybe available in a poster? (Woehl et al., 2007);
- The other is the Dolan (1997) publication about the UK tariff/weights of the EQ-5D
- One included results of TTO and VAS evaluations of utilities in 48 UC patients (steroid-refractory), Arseneau et al. (2006) – However, results reported on the 3 studies above were anyway different
- The last one was about a clinical activity index used to group patients by disease severity (Walmsley et al., 1998)
Adverse events, and in particular SAE are often taken into account into the HSUVs with different methodologies, like multiplicative or additive models.

However choice of the method is rarely justified and the overall calculation is not always transparent.

While the bibliographic source is often cited, the reporting of more than the point-estimate (average) of the disutility of the AE is rarely done.

Moreover the disutility of the AE might also come from publications on other disease areas than the one studied in the CEA model.
Challenges in reporting HSUVs – experiences from the industry

- Theory versus Practice of generating/using HSUVs for analytical decision models?
- How are HSUVs reported in models developed by manufacturers?
- Can we do better? Likely!
Search, review and appraise

Andrew Lloyd
Outline

• Limitations of RCT data for utilities
• Impact on SR methods
  • Data sources
  • Searching & selecting
  • Critical appraisal
Background to SR planning

• Clinical trials have limitations for collection of data
  • Very good at measuring specific treatment effect against a comparator

• Limitations for measuring HRQL
  • Placebo effects
    • Patients’ expectations regarding new treatment
    • Protocol driven interventions (different to routine practice)
    • Additional attention from clinical staff
  • Study entry criteria
    • Affects generalisability

• HRQL measures are subjective reports
  • Affected by subjective biases such as placebo effects
  • Trials not powered for collection of HRQL/ utility data
Placebo problem in RCTs

• Blue line shows data captured in trial
• Orange line = ‘true’ score
• Systematic error caused by
  • Placebo effects
  • Expectations
• Absolute vs Relative scores
• Reverse effect also possible in open label trials?
Avoiding bias – increasing generalisability

• Avoiding placebo effects
  • If participants have no expectation of improvement
  • Minimising additional clinical contact
  • Make patient experience as much like ‘routine care’ as possible

• Improve generalisability
  • Broad study entry criteria
  • Study sample to match model population
    • Care received/ nationality/ age/ comorbidities
Impact on SRs......

• Systematic reviews include
  1. Systematic & thorough search strategies
  2. Identification of target papers meeting criteria
  3. Critical appraisal of papers
  4. Accepted hierarchy of methods.... (....needed for utilities)

• For utilities
  • Need to go beyond RCTs for data
  • Databases/ observational research – more generalisable/ representative?
1. Systematic & thorough search strategies

- Must consider structure of cost effectiveness model
  - How are states defined; AEs; age; geography
- Search strategies similar to other systematic reviews
- Standard approach will identify published literature, but.....
- Studies need:
  - Clinical variable for defining patients into model states (e.g. NYHA)
  - Recognised utility measure – EQ-5D etc
- Other sources of utilities (Not on Medline etc)
  - Observational studies, Routine Outcome Measurement
  - Surveys
  - Databases (existing datasets) or ScHARR HUD/ Tufts
2. Identifying papers

- Target indication
  - Recency of study – has clinical practice moved on?
- Model states covered?
- Relevant HRQL measures?
- Relevant clinical measures?
3. Critical appraisal

1. Free from sources of methodological bias
   
   and

2. Meets the methodological standards of HTA body
   • Varies by jurisdiction

• Studies must meet both criteria
• Checklist
3a. Critical appraisal - bias

• Methodological bias
  • Study entry criteria/ selection bias
  • Regression to the mean; placebo effects
  • Non-random missing data
  • Inadequate sample size
  • Unrepresentative

• Methods of HRQL

• Clinical measures don’t align
3b. Critical appraisal – HTA needs

- Appropriate measure of HRQL (e.g. EQ-5D/ HUI-3)
- Appropriate national preference weights
- Mapping conducted to recognised standards
- Analysis of HRQL data
Conclusions

• Systematic reviews should be driven by
  • Model design/ structure
  • Needs of HTA body (e.g. NICE)

• Markers of quality
  • Reflect decision problem
  • Data that are most suited or informative for decision
  • Free from measurement bias
Synthesising HSUV

Roberta Ara, University of Sheffield
HSUV Taskforce, November 2017
“Researchers wishing to populate decision analytic models have a responsibility to incorporate all high-quality evidence available” [Peasgood 2015]

Should we synthesis HSUVs?

Outline:
• Decision analytic models
• International requirements
• QoL measures & methods
• Studies that have synthesised HSUVs
Decision Analytic Models – Markov models
<table>
<thead>
<tr>
<th>Country</th>
<th>Preferred instrument</th>
<th>Alternative instrument</th>
<th>Direct assessment</th>
<th>Preference weights</th>
<th>Elicitation method</th>
<th>Who report QoL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>GPBM</td>
<td>Direct, mapped</td>
<td>yes</td>
<td>own public</td>
<td>SG or TTO</td>
<td>Patient</td>
</tr>
<tr>
<td>Canada</td>
<td>GPBM direct utility assessment</td>
<td>Willingness to pay</td>
<td>yes</td>
<td>any public</td>
<td>SG or TTO</td>
<td>Patient</td>
</tr>
<tr>
<td>France</td>
<td>EQ-5D/HUI3 (other GPBM)</td>
<td>CSPBM, direct valuation from specific questionnaire</td>
<td>-</td>
<td>own public</td>
<td>SG or TTO</td>
<td>Patient #</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>EQ-5D-5L</td>
<td>EQ-5D-3L, other GPBM, CSPBM, mapped, direct utility assessment</td>
<td>-</td>
<td>own public</td>
<td>SG or TTO</td>
<td>Patient</td>
</tr>
<tr>
<td>Spain</td>
<td>EQ-5D &amp; SF-6D</td>
<td>other GPBM</td>
<td>-</td>
<td>own public</td>
<td>SG or TTO</td>
<td>Patient #</td>
</tr>
<tr>
<td>Sweden</td>
<td>direct utility assessment</td>
<td>PBM (eg EQ-5D)</td>
<td>yes</td>
<td>patients</td>
<td>SG or TTO</td>
<td>Patient</td>
</tr>
<tr>
<td>UK NICE</td>
<td>EQ-5D</td>
<td>mapped values, other measures</td>
<td>no</td>
<td>own public</td>
<td>choice based</td>
<td>Patient #</td>
</tr>
<tr>
<td>UK Scotland</td>
<td>EQ-5D</td>
<td>mapped values, direct utility assessment</td>
<td>yes</td>
<td>public</td>
<td>SG or TTO</td>
<td>Patient</td>
</tr>
</tbody>
</table>

# if infeasible or inappropriate, allow proxy
Measures and methods used to quantify utility

- **GPBM**
  - EQ-5D, SF-6D, HUI etc

- **CSPBM (~30)**
  - EORTC, HAQ-PBM, ReQoL etc

- **Mapped values**
  - between QoL
  - from clinical measure

- **Preference weights**
  - Country specific
  - Public, patient, expert

- **Elicitation technique**
  - TTO, SG, judgement, direct rating

- **Mode of administration**
  - Paper, electronic

- **Who completes**
  - Patient, proxy

- **Vas**

Utility
- 1 = full health
- 0 = death
- -ve = worse than death
## Variation across GPBM ranges

<table>
<thead>
<tr>
<th>Instrument</th>
<th>HSUV range</th>
<th>Country</th>
<th>Valuation technique</th>
<th>Model type</th>
</tr>
</thead>
<tbody>
<tr>
<td>15D</td>
<td>0.11 - 1</td>
<td>Finland</td>
<td>VAS</td>
<td>MAUT additive</td>
</tr>
<tr>
<td>AQoL-8D</td>
<td>-0.04 - 1</td>
<td>Australia</td>
<td>VAS transformed into TTO</td>
<td>MAUT multiplicative &amp; statistical</td>
</tr>
<tr>
<td>EQ-5D-3L</td>
<td>-0.xx - 1</td>
<td>numerous</td>
<td>ranking, TTO, VAS</td>
<td>Statistical additive</td>
</tr>
<tr>
<td>HUI3</td>
<td>-0.36 - 1</td>
<td>Canada, France</td>
<td>VAS transformed into SG</td>
<td>MAUT multiplicative</td>
</tr>
<tr>
<td>SF-6D</td>
<td>0.30 - 1</td>
<td>numerous</td>
<td>SG, ranking</td>
<td>Statistical additive with interaction</td>
</tr>
<tr>
<td>QWB-SA</td>
<td>0.08 - 1</td>
<td>USA</td>
<td>VAS</td>
<td>Statistical additive</td>
</tr>
</tbody>
</table>
EQ-5D-3L variation across setting [Janssen, 2014]
Variation in HSUVs obtained using different instruments [Brazier 2004]

<table>
<thead>
<tr>
<th></th>
<th>EQ-5D</th>
<th></th>
<th>SF-6D</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>sd</td>
<td>mean</td>
<td>sd</td>
</tr>
<tr>
<td>Lower back pain</td>
<td>0.614</td>
<td>0.299</td>
<td>0.662</td>
<td>0.141</td>
</tr>
<tr>
<td>CPD</td>
<td>0.540</td>
<td>0.309</td>
<td>0.572</td>
<td>0.112</td>
</tr>
<tr>
<td>IBS</td>
<td>0.662</td>
<td>0.260</td>
<td>0.666</td>
<td>0.146</td>
</tr>
<tr>
<td>Leg ulcer</td>
<td>0.636</td>
<td>0.266</td>
<td>0.658</td>
<td>0.144</td>
</tr>
<tr>
<td>Osteoarthritis</td>
<td>0.552</td>
<td>0.397</td>
<td>0.647</td>
<td>0.145</td>
</tr>
<tr>
<td>Over 75 years</td>
<td>0.729</td>
<td>0.262</td>
<td>0.716</td>
<td>0.143</td>
</tr>
<tr>
<td>Menopausal women</td>
<td>0.442</td>
<td>0.336</td>
<td>0.521</td>
<td>0.114</td>
</tr>
</tbody>
</table>
Should we synthesise HSUVs?

Methods used to synthesise the evidence

• Meta-analysis
  – provides a weighted point estimate, increase power & precision
  – does not take into account important differences, methods used & population

• Meta-regression
  – explore differences caused by variation in study design, methods etc.
  – needs many more data points (min 10 recommended per covariate)
<table>
<thead>
<tr>
<th>Condition</th>
<th>Studies (N)</th>
<th>HSUVs (N)</th>
<th>Measures/techniques</th>
<th>Countries (N)</th>
<th>meta-analysis (I²)</th>
<th>meta-regress</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>CKD</td>
<td>190</td>
<td>326</td>
<td>SF-36, SF12, 15D, SF-6D, EQ-5D, HUI, TTO, SG, mapped values</td>
<td>multiple</td>
<td>-</td>
<td>y</td>
<td>Wyld, 2012</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>26</td>
<td>351</td>
<td>SG, TTO, EQ-5D, HUI3, VAS</td>
<td>7</td>
<td>-</td>
<td>y</td>
<td>Djalalov, 2014</td>
</tr>
<tr>
<td>CHD</td>
<td>40</td>
<td>&gt;80</td>
<td>EQ-5D, 15D, QWB, SF-6D, HUI, TTO, RS, HALex</td>
<td>&gt;15</td>
<td>&gt;0.71</td>
<td>y</td>
<td>Stevanovic, 2016</td>
</tr>
<tr>
<td>Diabetes</td>
<td>45</td>
<td>66</td>
<td>EQ-5D, HUI3, SF-6D, TTO, SG</td>
<td>NR</td>
<td>0.98</td>
<td>y</td>
<td>Lung, 2011</td>
</tr>
<tr>
<td>Liver disease</td>
<td>6</td>
<td>40</td>
<td>VAS, HUI2, HUI3, TTO, SG, EQ-5D, TVAS, AQOL, judgement (delphi techniques)</td>
<td>-</td>
<td>-</td>
<td>y</td>
<td>Mclernon, 2008</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>24</td>
<td>223</td>
<td>SG, judgement, direct rating, HALex, AQOL, EQ-5D, TTO</td>
<td>7</td>
<td>-</td>
<td>y</td>
<td>Sturza, 2010</td>
</tr>
<tr>
<td>Osteoporosis</td>
<td>62</td>
<td>362</td>
<td>EQ-5D, VAS, SG, TTO, HUI, SF-36, QWB</td>
<td>multiple</td>
<td>0.99 to 1</td>
<td>y</td>
<td>Si, 2014</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>23</td>
<td>173</td>
<td>TTO, SG, judgment, Rating scale, QWB, HUI</td>
<td>NR</td>
<td>-</td>
<td>y</td>
<td>Bremner, 2007</td>
</tr>
</tbody>
</table>
Existing studies (2)

Huge amounts of heterogeneity across instruments & studies:

‘analysts should avoid direct comparisons of lung cancer utility values elicited with dissimilar methods’

‘caution when comparing values across instruments’

‘this heterogeneity limits the meaningfulness of statistical pooling’

‘uncertainty is considerable and is mostly found between studies’

‘provides a standard set of HSUVs that can be used in health economic assessments’
Should we synthesise HSUVs?

Difference in HSUVs instruments/methods
  GPBM/CSPBM
Techniques used to elicit weights
Mode of collection
Who completes the questionnaires
Preference weights

Statistical techniques
References


Peasgood T, Brazier J. Is Meta-analysis for utility values appropriate given the potential impact different elicitation methods have on values? Pharmacoeconomics (2015) 33:1101-1105


Poll: Should exhaustive searches be undertaken to identify HSU estimates for cost-effectiveness models?
Poll: Should we be synthesising HSU estimates for cost-effectiveness models?
Poll: Do HSU estimates from RCTs constitute the best evidence for cost-effectiveness models?
Pre/Post Comparison: Do HSU estimates from RCTs constitute the best evidence for cost-effectiveness models?
Next stages....

• Further iteration of HSU Estimates Taskforce Good Practice development
• Review of draft HSU Estimates Good Practice
• Publication 2018
• Taskforce Workforce / Short Course