ISPOR Taskforce: Identification of evidence and use in Cost-Effectiveness Models: Good Practices for Outcomes Research

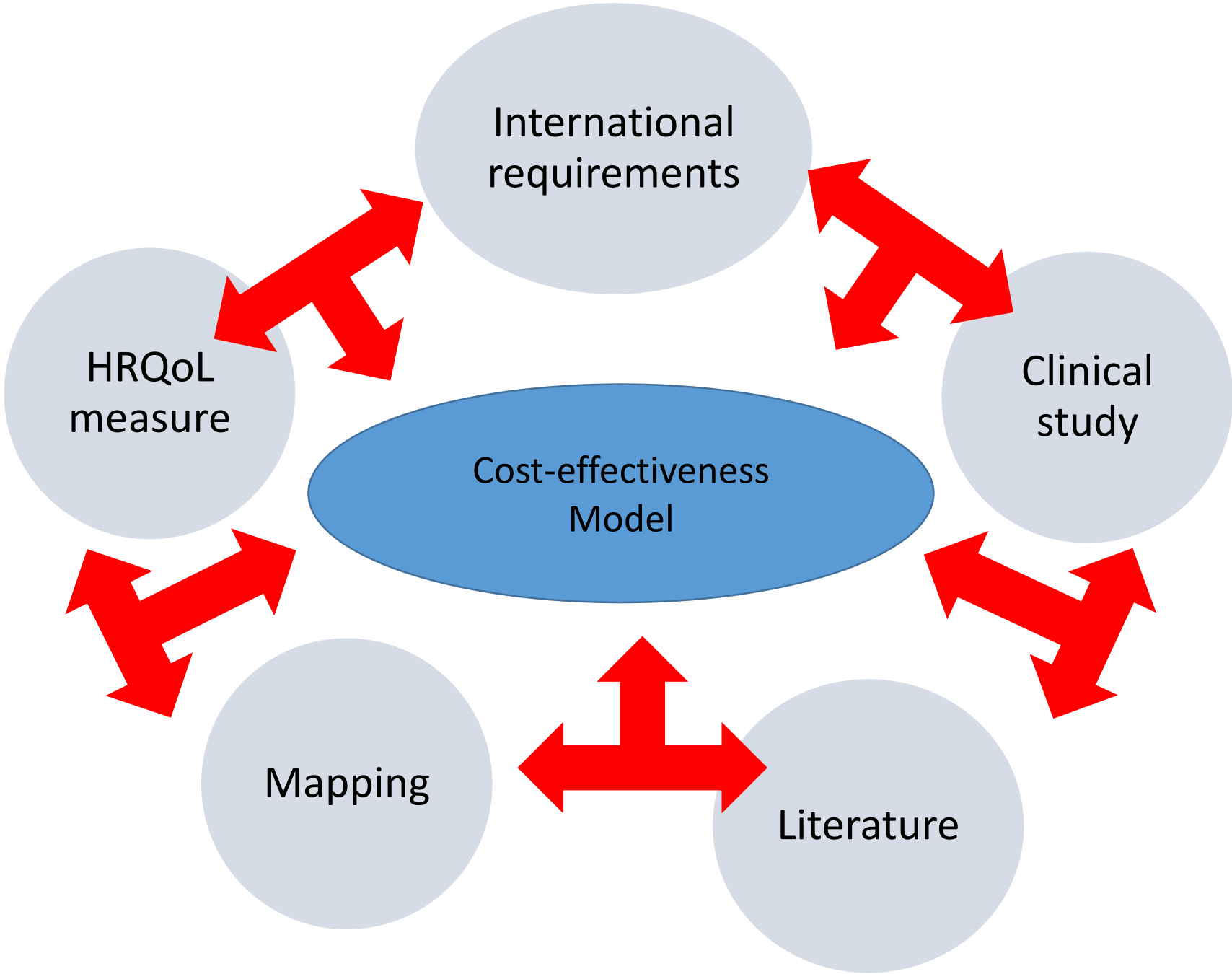
Taskforce Representatives:

Suzy Paisley (Moderator)

Hélène Chevrou-Séverac

Andrew Lloyd

Roberta Ara


International requirements
HRQoL measure
Clinical study
Cost-effectiveness Model
Mapping
Literature

ISPOR Task Force: Identification of evidence and use in Cost-Effectiveness Models: Good Practices for Outcomes Research

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

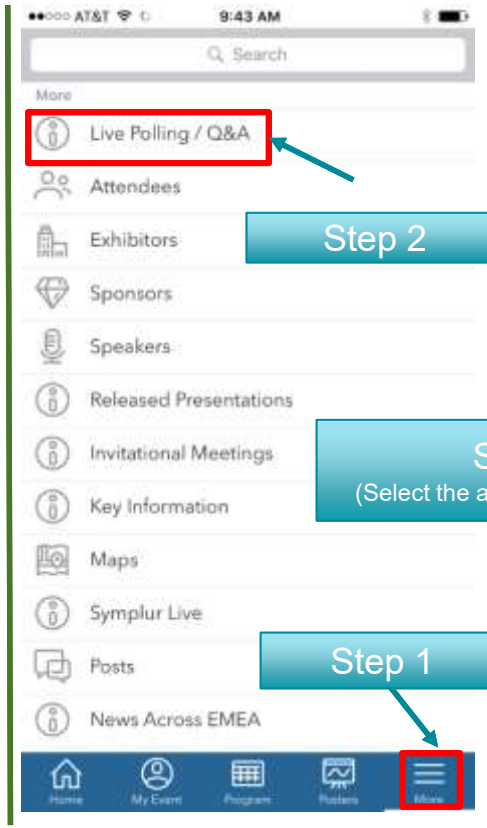
WiFi Network: ISPORGlasgow | Password: IQVIA2017

https://myispor.cnf.io/

ISPOR **Congress Platform**

Mobile App

Search "**ISPOR**"
in the App Store or on
Google Play!



F7: HEALTH STATE UTILITY (HSU) IDENTIFICATION AND USE IN COST-EFFECTIVENESS IN DECISION MODELLING - AN ISPOR TASK FORCE

Live Content Slide

When playing as a slideshow, this slide will display live content

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?

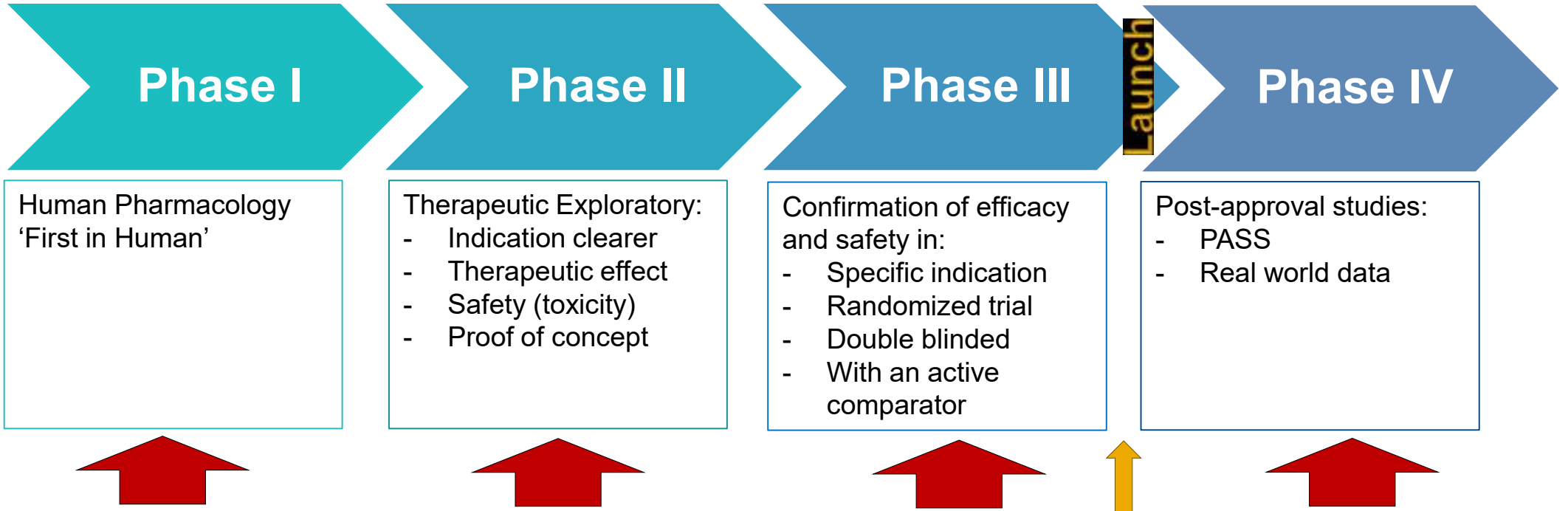
Celgene

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?

Celgene

Theory of generating and reporting HSUVs



Celgene

Some practices in generating and reporting HSUVs

Phase I

Human Pharmacology
'First in Human'

Some exploration of the disease and treatment pathway - *not always involvement of HEOR/PMA*

Phase II

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

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

Phase III

Launch

Confirmation of efficacy and safety in
- Specific i
- Randomiz
- Double bl
- With an a comparat

No always HR-QoL instruments included, or HSUV, not often in the lon term; not always knowledg about how to analyze ther

Phase IV

Post-approval studies:

orld data

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

Celgene

Missing HSUVs from het phase 3 randomized clinical trial program

- Cases whe
  - Start-ups … he time of the develop…
  - When the … ssions; and the RCT from the … after 5-10 years)
  - When it …

How is it reported by the industry in CEA publications?

- 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

Celgene

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?

Celgene

Justifying model choice and inputted data is key in HTA submissions

"The rig *defined criteria b*

justifying *n decisions"*

How good are we in the HEOR/HTA field and Medical industry in doing so?

Paisley, S. "Identification of Evidence for Key Parameters in Decision-Analytic Models of Cost-Effectiveness", 2016 PharmacoEconomics, 34(6), pp. 597-608.

Celgene

Reporting of HSUVs from the RCT data for use in CEA models

- 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)

Celgene

Reporting of HSUVs extracted from a *Systematic* Literature Search

- 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.

« NICE DSU TECHNICAL SUPPORT DOCUMENT 13: IDENTIFYING AND REVIEWING EVIDENCE TO INFORM THE CONCEPTUALISATION AND POPULATION OF COST-EFFECTIVENESS MODELS", Kaltenthaler et al. 2011

Reporting of HSUVs extracted from literature searches

- 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)

Health state costs and utility weights	Cost/utility by health state					
Parameter	**Remission**	**Mild**	**Moderate-severe**	**Surgery**	**Post-surgery remission**	**Post-surgery complications**
Per-cycle cost	£240.49	£431.05	£973.86	£13,880.32[d]	£475.50	£1,945.38
Per-cycle utility weight	0.88	0.76	0.42	0.42	0.60	0.42

  - "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."

Celgene

Reporting of HSUVs extracted from literature searches

The Russian Nesting Dolls-reporting approach

- **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 referen**
  - **Tsay et al. (2008)** :"
  - **Punekar et al. (201)**
- **Initial publications u**
  - One is an abstract o
  - The other is the Dola
  - One included results
  - The last one was about a clinical activity index used to group patients by disease severity (Walmsley et al., 1998)

his benefit was quantified using published data
nt survey carried out in Cardiff Hospital using the E
of the UK population.13, 17"
ptom-free remission and surgical complicatior
e health state preferences obtained from a UC
5D [16] and valued using UK tariffs [17]."
ct doesn't include any Utilities data– maybe availab
-5D
(steroid-refractory), Arseneau et al. (2006) – However, results reported on the 3 studies above were anyway different

Reporting of ‘adaptation’ of HSUVs for Adverse Events

- 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

Celgene

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 !

Celgene

Search, review and appraise

Andrew Lloyd

Outline

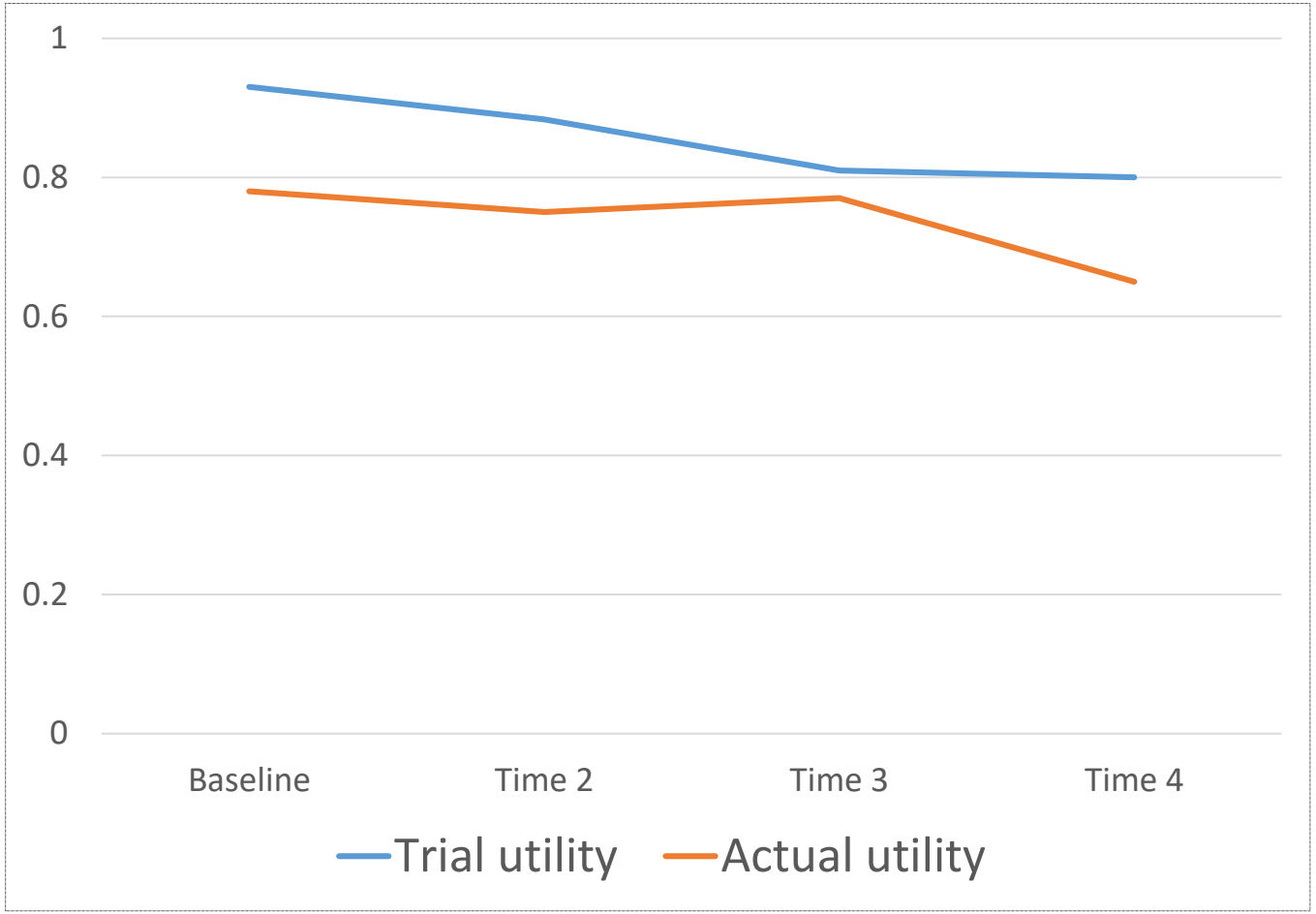
- Limitations of RCT data for utilities
- Impact on SR methods
  - Data sources
  - Searching & selecting
  - Critical appraisal

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

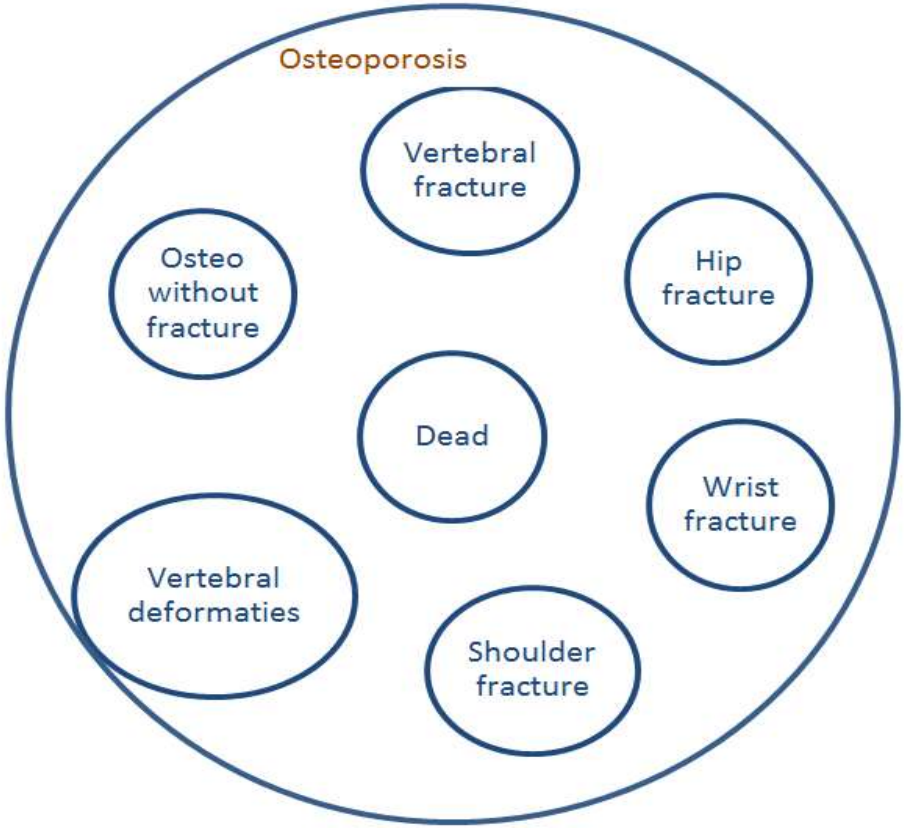
"Researchers wishing to populate decision analytic models have a responsibility to incorporate all high-quality evidence available" [Peasgood 2015]

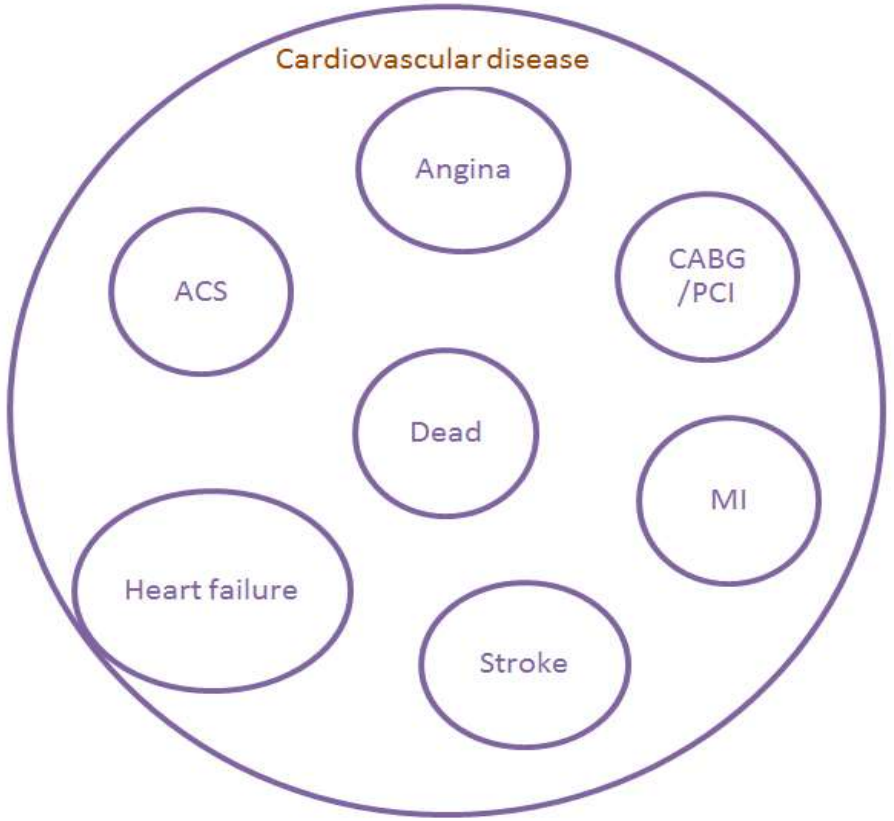
Should we synthesis HSUVs?

Outline:

- Decision analytic models
- International requirements
- QoL measures & methods
- Studies that have synthesised HSUVs

Decision Analytic Models – Markov models





International recommendations [Rowen et al , 2017]

	Preferred instrument	Alternative instrument	Direct assessment	Preference weights	Elicitation method	Who report QoL
Australia	GPBM	Direct, mapped	yes	own public	SG or TTO	Patient
Canada	GPBM direct utility assessment	Willingness to pay	yes	*any public*	SG or TTO	Patient
France	EQ-5D/HUI3 (other GPBM)	CSPBM, direct valuation from specific questionnaire	-	own public	SG or TTO	Patient[#]
The Netherlands	EQ-5D-5L	EQ-5D-3L, other GPBM, CSPBM, mapped, direct utility assessment	-	own public	SG or TTO	Patient
Spain	EQ-5D & SF-6D	other GPBM	-	own public	SG or TTO	Patient[#]
Sweden	direct utility assessment	PBM (eg EQ-5D)	yes	**patients**	SG or TTO (& VAS)	Patient
UK NICE	EQ-5D	mapped values, other measures	no	own public	choice based	Patient[#]
UK Scotland	EQ-5D	mapped values, direct utility assessment	yes	public	SG or TTO	Patient

[#] 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

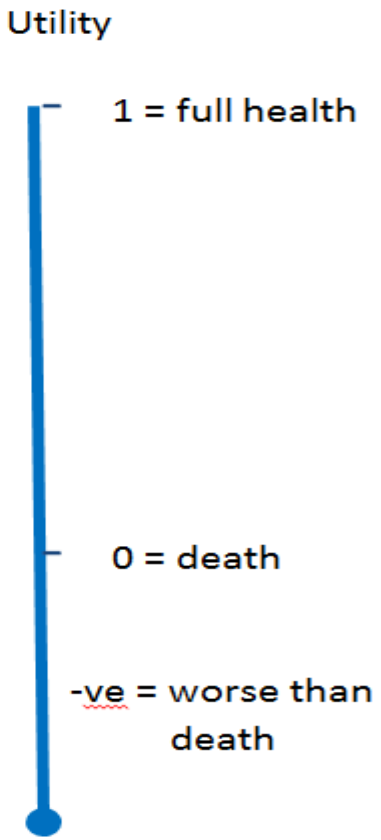
Preference weights
- Country specific
- Public, patient, expert

Elicitation technique
- TTO, SG, judgement, direct rating

Vas

Mode of administration
- paper, electronic

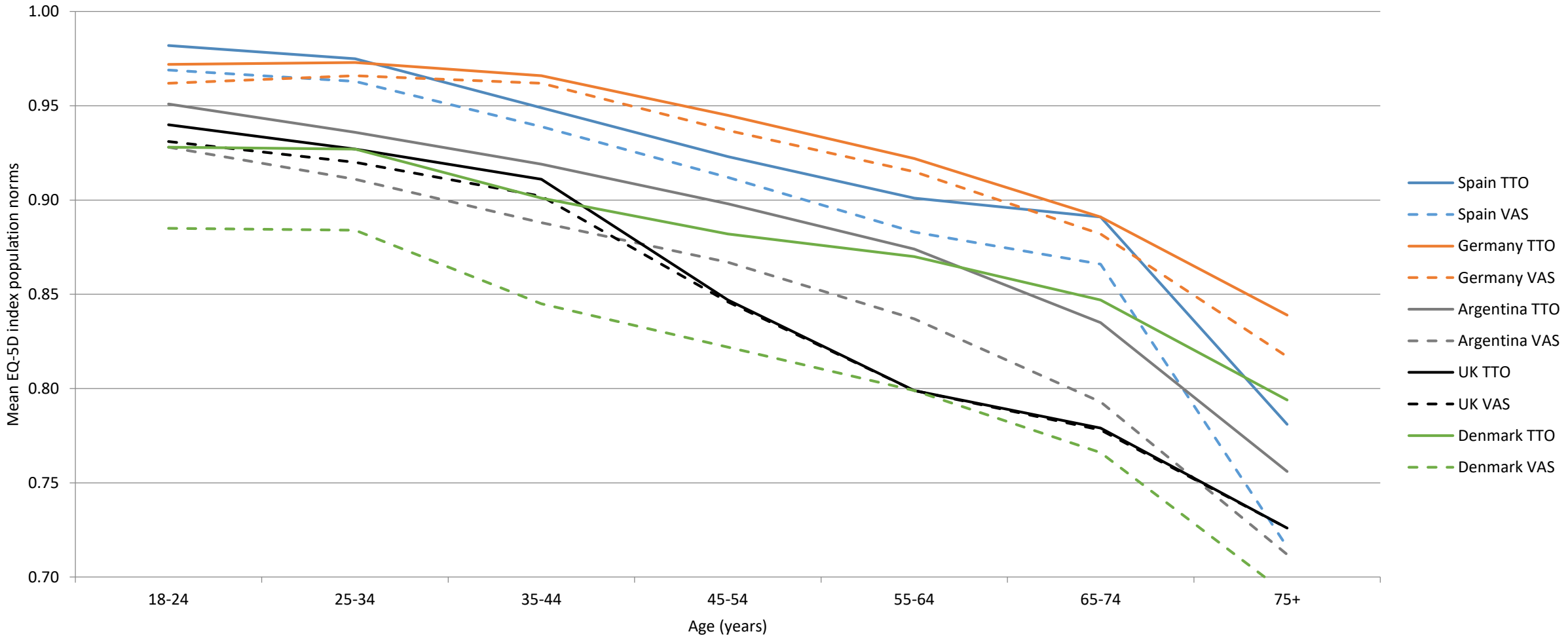
Who completes
- Patient, proxy



Variation across GPBM ranges

Instrument	HSUV range	Country	Valuation technique	Model type
15D	0.11 - 1	Finland	VAS	MAUT additive
AQoL-8D	-0.04 - 1	Australia	VAS transformed into TTO	MAUT multiplicative & statistical
EQ-5D-3L	-0.xx - 1	numerous	ranking, TTO, VAS	Statistical additive
HUI3	-0.36 - 1	Canada, France	VAS transformed into SG	MAUT multiplicative
SF-6D	0.30 - 1	numerous	SG, ranking	Statistical additive with interaction
QWB-SA	0.08 - 1	USA	VAS	Statistical additive

EQ-5D-3L variation across setting [Janssen, 2014]



Variation in HSUVs obtained using different instruments [Brazier 2004]

	EQ-5D		SF-6D	
	mean	sd	mean	sd
Lower back pain	0.614	0.299	0.662	0.141
CPD	0.540	0.309	0.572	0.112
IBS	0.662	0.260	0.666	0.146
Leg ulcer	0.636	0.266	0.658	0.144
Osteoarthritis	0.552	0.397	0.647	0.145
Over 75 years	0.729	0.262	0.716	0.143
Menopausal women	0.442	0.336	0.521	0.114

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)

Exemplars of existing studies that synthesis HSUVs [Peasgood 2015]

Condition	Studies (N)	HSUVs (N)	Measures/techniques	Countries (N)	meta-analysis (I^2)	meta-regress	Ref
CKD	190	326	SF-36, SF12, 15D, SF-6D, EQ-5D, HUI, TTO, SG, mapped values	multiple	-	y	Wyld, 2012
Colorectal cancer	26	351	SG, TTO, EQ-5D, HUI3, VAS	7	-	y	Djalalov , 2014
CHD	40	>80	EQ-5D, 15D, QWB, SF-6D, HUI, TTO, RS, HALex	>15	>0.71	y	Stevanovic, 2016
Diabetes	45	66	EQ-5D, HUI3, SF-6D, TTO, SG	NR	0.98	y	Lung, 2011
Liver disease	6	40	VAS, HUI2, HUI3, TTO, SG, EQ-5D, TVAS, AQOL, judgement (delphi techniques)	4	-	y	Mclernon, 2008
Lung cancer	24	223	SG, judgement, direct rating, HALex, AQOL, EQ-5D, TTO	7	-	y	Sturza, 2010
Osteoperosis	62	362	EQ-5D, VAS, SG, TTO, HUI, SF-36, QWB	multiple	0.99 to 1	y	Si, 2014
Prostate cancer	23	173	TTO, SG, judgment, Rating scale, QWB, HUI	NR	-	y	Bremner, 2007

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

Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health economics. 2004 Sep 1;13(9):873-84.

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

Rowen D, Azzabi Zouraq I, Chevrou-Severac H, van Hout B. International Regulations and Recommendations for Utility Data for Health Technology Assessment. Pharmacoeconomics. 2017 https://link.springer.com/article/10.1007%2Fs40273-017-0544-y

Janssen B., Szende A. (2014) Population Norms for the EQ-5D. In: Szende A., Janssen B., Cabases J. (eds) Self-Reported Population Health: An International Perspective based on EQ-5D. Springer, Dordrecht

Live Content Slide

When playing as a slideshow, this slide will display live content

Poll: Should exhaustive searches be undertaken to identify HSU estimates for cost-effectiveness models?

Live Content Slide

When playing as a slideshow, this slide will display live content

Poll: Should we be synthesising HSU estimates for cost-effectiveness models?

Live Content Slide

When playing as a slideshow, this slide will display live content

Poll: Do HSU estimates from RCTs constitute the best evidence for cost-effectiveness models?

Live Content Slide

When playing as a slideshow, this slide will display live content

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