RTI HEALTH SOLUTIONS®

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

Educational Symposium: 3 June 2014

RTI Health Solutions

- Research Triangle Park, NC, USA
- Ann Arbor, MI, USA
- Barcelona, Spain
- Ljungskile, Sweden
- Manchester, UK
- Waltham, MA, USA

www.rtihs.org
e-mail: rtihealthsolutions@rti.org

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 co-morbidities 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.

Measuring utility for economic models within clinical trials: can we do better?

MODERATOR	**Sorrel Wolowacz, PhD**, Head, Health Economics Europe, RTI Health Solutions, UK
PANEL	**Jennifer Petrillo, PhD**, Director, Global Health Economics & Outcomes Research, Novartis AG, Switzerland **Lynda Doward, MRes**, European Head of Patient-Reported Outcomes, RTI Health Solutions, UK **Andrew Briggs, DPhil**, Professor, Health Economics and Health Technology Assessment, University of Glasgow, UK

Issues with utility data collected in trials

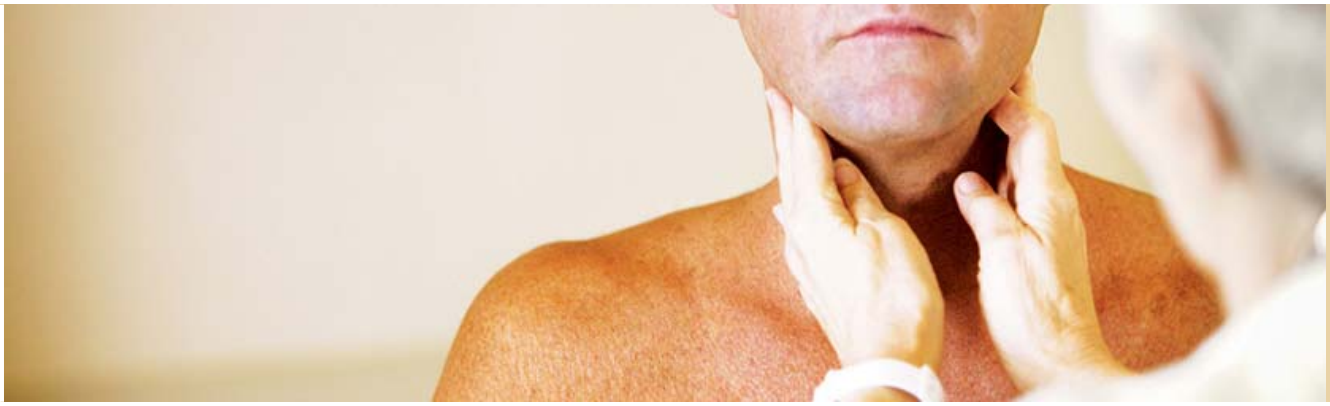
- EQ-5D data are commonly being collected alongside clinical trials
  - Primary purpose is to provide utility weights for economic models
  - In practice, data collected has often had severe limitations
- Issues with trial design
  - Limitations in patient follow-up often results in key health states 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 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
  - Input to scope of Good Research Practices Task Force (proposal currently being considered by ISPOR Health Science Policy Council)

Panel perspectives

Jennifer Petrillo **Novartis**	Industry perspective - health economics & outcomes research
Lynda Doward **RTI-HS**	Patient-reported outcomes research
Andy Briggs **Glasgow University**	Health economic modelling



Issues With Utility Data: Industry Perspective

Jennifer Petrillo, PhD

3 June 2014

NOVARTIS

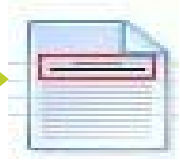
Why are utilities so important..

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)

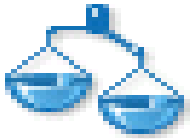
Step 3
Submit ICER evidence to reimbursement bodies for acceptance of product

Intervention administered

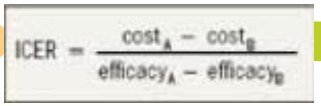


Patients complete EQ-5D to reflect utility

Average trial sample ~300 patients



Utilities input to cost-effectiveness model

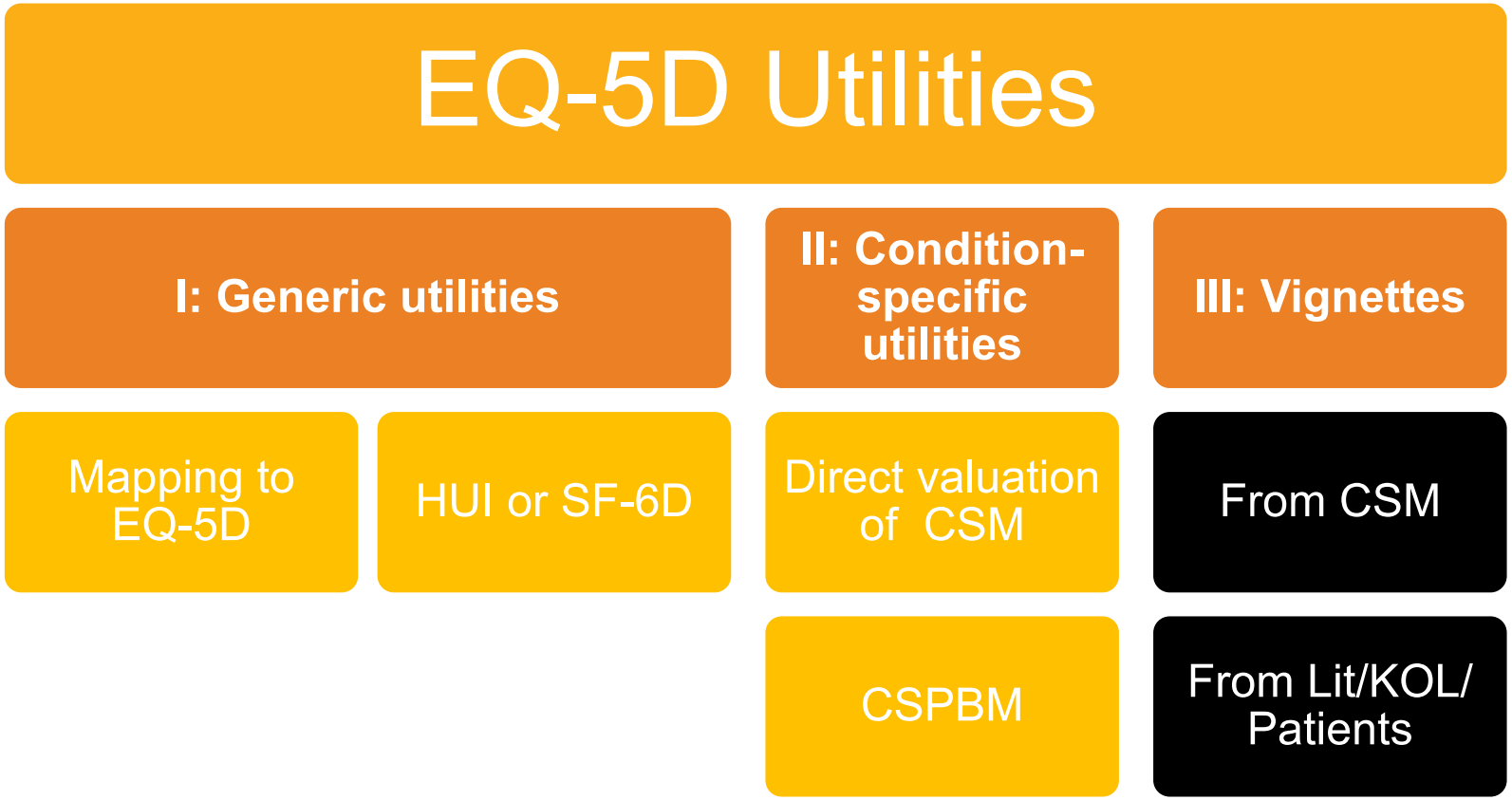


ICER

(hopefully) NICE Decision

NOVARTIS

NICE (P)reference Case: a hierarchy of approaches



NOVARTIS

How utilities are currently collected: Example 1

- Phase II dose ranging study in PD-LID
  - Assessment at Baseline and Wk9
  - Modified Abnormal Involuntary Movement Scale (mAIMS)
  - EQ-5D

Figure 1. Plot of change in EQ-5D utility score between baseline and week 9 by change in mAIMS total score between baseline and week 9

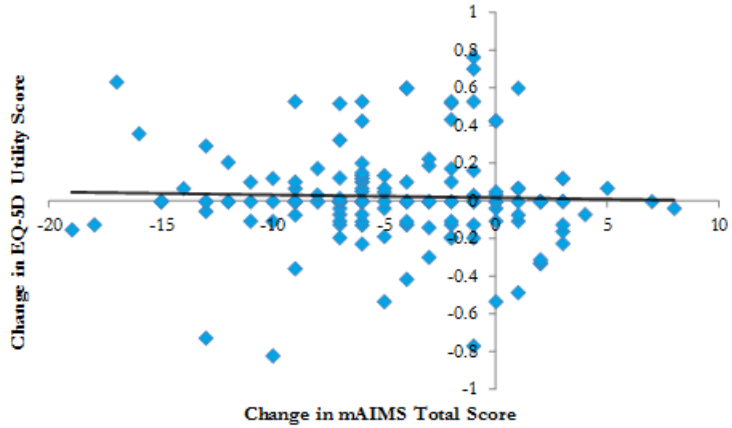


Table 2. Change in mAIMS, EQ-5D utility, and EQ-5D VAS scores between Baseline and Week 9

Measure	Mean (SD)	Min	Q1	Median	Q3	Max
Δ mAIMS	-4.8 (5.2)	0	2	5	8	19
Δ EQ-5D utility	0.024 (0.255)	-0.820	-0.104	0.000	0.104	0.768
Δ EQ-5D VAS	-0.05 (16.07)	-45	-10	0	10	41

- 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).

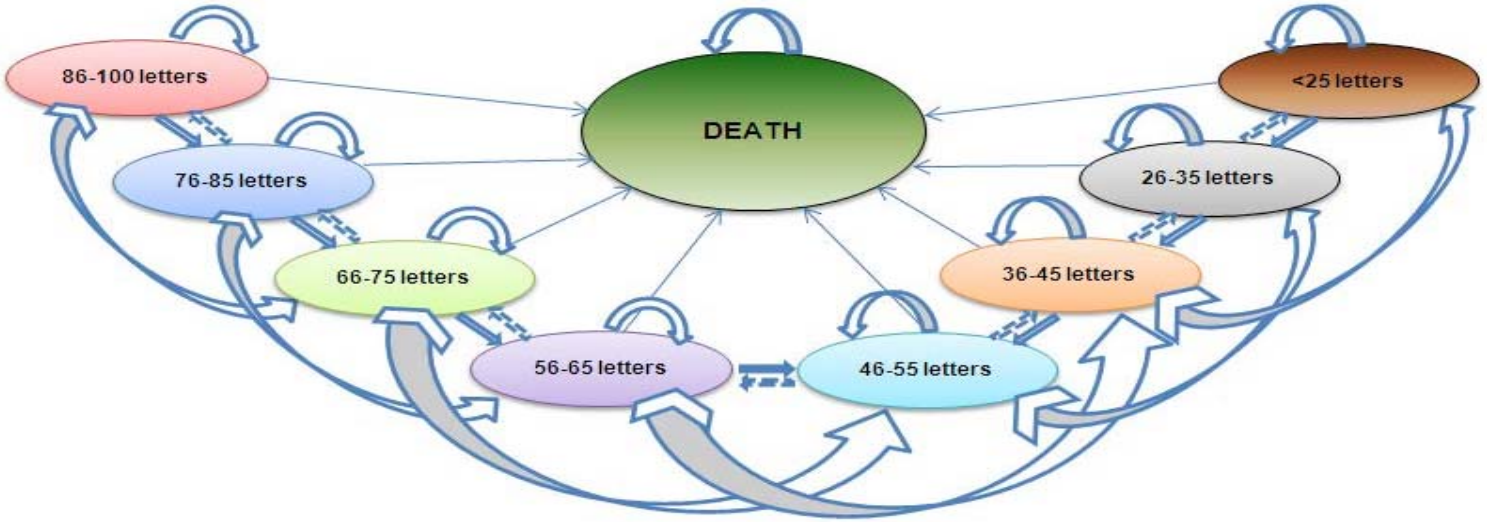
- Issues: Small sample, uncertain dose/treatment regimen, not powered for differences, PD-LID concepts may be unrelated to EQ-5D

NOVARTIS

Presented at ISPOR 18th Annual International Meeting, May 18-22, 2013, New Orleans, LA.

How utilities are currently collected: Example 2

- Phase 3 study in Diabetic Macular Edema
  - Visit dates at Baseline, 3, and 12 months
  - EQ5D
  - VFQ-25



Ref: Assessment of utility loss from diabetic macular edema based on RESTORE trial. M Knudsen, S Thomas, P Mitchell. Association for Research in Vision and Ophthalmology (ARVO), May 6–10, 2012, Florida, USA.

NOVARTIS

How utilities are currently collected: Example 2

Table 1 Utility by BCVA in treated eye

Health state defined by BCVA category (letters; treated eye)	RESTORE*	Lloyd *et al*†	WSE RESTORE	BSE RESTORE
	Mean utility (SE)			
1: 86–100	0.860‡ (0.034)	0.830		
2: 76–85	0.860 (0.014)	0.750		
3: 66–75	0.813 (0.012)	0.750		
4: 56–65	0.802 (0.014)	0.715		
5: 46–55	0.770 (0.018)	0.680		
6: 36–45	0.760 (0.027)	0.680	0.728 (0.044)	0.698 (0.073)
7: 26–35	0.681 (0.053)	0.530	0.487 (0.088)	na.
8: 0–25	0.547 (0.083)	0.340	0.785 (0.115)	na.

*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 measurement points to be able to cover all possible health state transitio in the pooled data was rejected ($p<0.05$).

†Patients underwent a Snellen visual acuity (VA) assessment and were adjustments were made to published values in order to convert VA ranges model.

‡Restricted to being greater than the utility in health state 2.

**86–100 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.

- It is difficult to make firm conclusions of the slope of the utility function because the steepest part is in the range where we have fewer or no observations. It is however worthwhile noting that when better seeing and worse seeing eyes are analysed separately we observe comparable slopes in the range from 75–100 to 36–45 letters. Due to missing observations of BSE we are unable to tell the difference between BSE and WSE in the range <36 letters.

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

Ref: Assessment of utility loss from diabetic macular edema based on RESTORE trial. M Knudsen, S Thomas, P Mitchell. Association for Research in Vision and Ophthalmology (ARVO), May 6–10, 2012, Florida, USA.

NOVARTIS

Limitations across trials

WHO

- Trial population may not be representative of patients in routine clinical practice
- Severity ranges, co-morbidities, 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)

RTI HEALTH SOLUTIONS®

MEASURING UTILITY
The Patient Perspective

Lynda C Doward, MRes

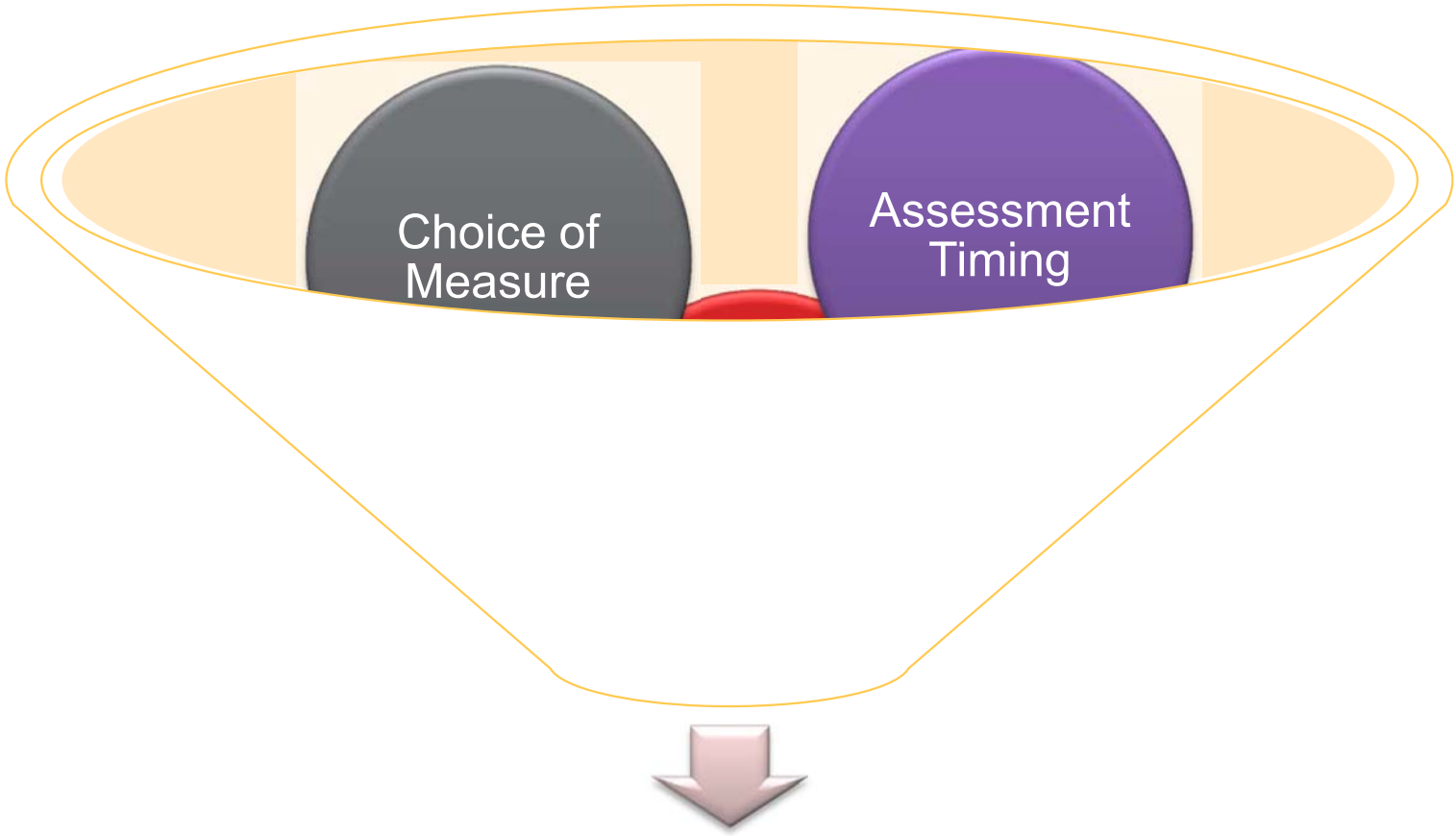
ISPOR Educational Symposium: 3 June 2014

RTI Health Solutions

- Research Triangle Park, NC, USA
- Ann Arbor, MI, USA
- Barcelona, Spain
- Ljungskile, Sweden
- Manchester, UK
- Waltham, MA, USA

www.rtihs.org
e-mail: rtihealthsolutions@rti.org

Taking the patient perspective …



Quality of Utility Data

Are we doing the best we can?

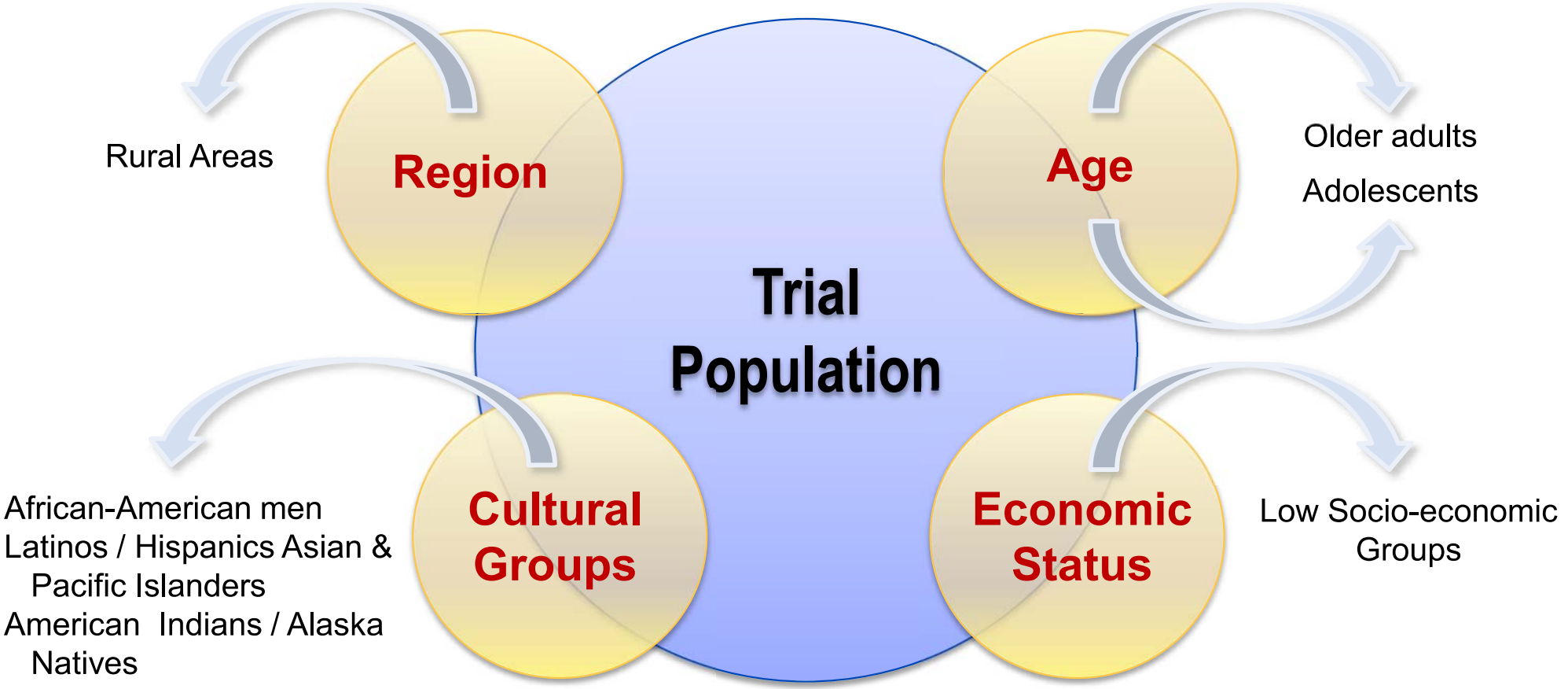
Are trial populations representative?

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

Age	Diagnostic sub-group	Disease severity	Co-morbidities
Geographic location	Socio-economic status	Race / ethnicity	Language

Underrepresentation in trials

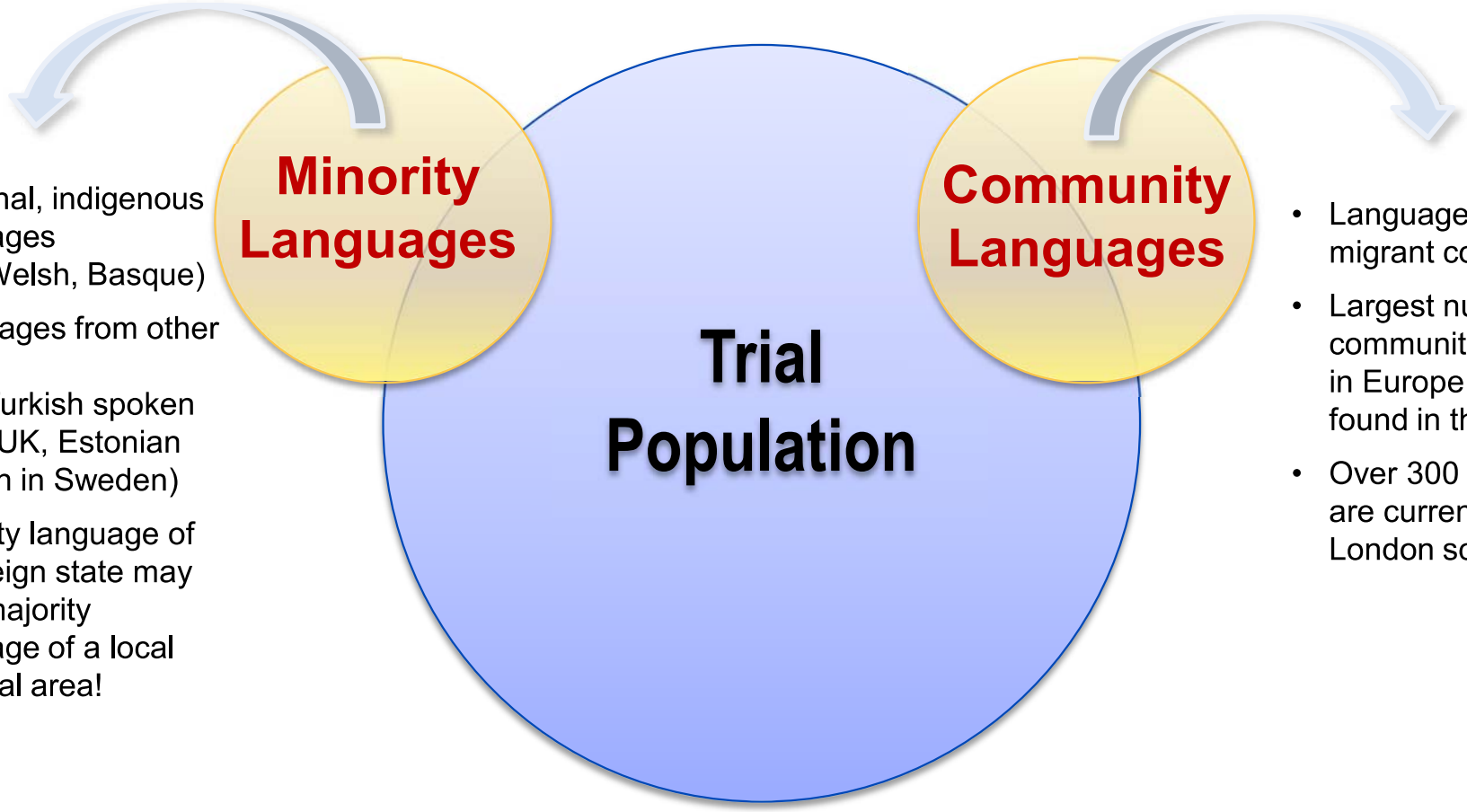
Only 2.5% of cancer patients are enrolled into NCI-funded clinical trials (Ford et. al., 2007)



Language

Trial populations tend to include the official or majority languages

- Regional, indigenous languages (e.g. Welsh, 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!



- 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 QALY’s 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
  - Content validity
    - 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 (RRMS)	• 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
COPD	• 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?

Are there alternatives?

- Key Question
  - How do we ensure that the utility values we use in economic modelling are the most realistic and representative?
- Can we optimise collection of utility data in RCTs?
  - Optimum trial design
  - Mapping non-preference condition-specific measures
- Alternatives to RCTs?
  - Observational studies
  - Surveys designed to collect utilities
  - Valuation of health state vignettes (time trade-off etc.)
- Collect empirical evidence for alternative methods!

Maximise use of mobile technology!

References

- Chiò A, Canosa A, Gallo S, Cammarosano S, Moglia C, Fuda G, Calvo A, Mora G; PARALS group. ALS clinical trials: do enrolled patients accurately represent the ALS population? Neurology. 2011 Oct 11;77(15):1432-7. doi: 10.1212/WNL.0b013e318232ab9b. Epub 2011 Sep 28 (Accessed 24 October 2013).
- European Medicines Agency (EMA). Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products. 2005. Available at: http://www.ispor.org/workpaper/emea-hrql-guidance.pdf. Accessed 31 January 2012.
- Fernández-Megía MJ, Casanova B, Magraner MJ, Font-Noguera I, Poveda-Andrés JL. Assessment of the effectiveness and safety of natalizumab for treating relapsing-remitting multiple sclerosis. Farm Hosp. 2011 Mar-Apr;35(2):75-9. doi: 10.1016/j.farma.2010.02.003. Epub 2010Aug 3.
- Ford JG, Howerton MW, Lai GY, Gary TL, Bolen S, Gibbons MC, Tilburt J, Baffi C, Tanpitukpongse TP, Wilson RF, Powe NR, Bass EB. Barriers to recruiting underrepresented populations to cancer clinical trials: a systematic review. Cancer. 2008 Jan 15;112(2):228-42. Review.
- Kessler R, Ståhl E, Vogelmeier C, Haughney J, Trudeau E, Löfdahl CG, Partridge MR. Patient understanding, detection, and experience of COPD exacerbations: an observational, interview-based study. Chest. 2006 Jul;130(1):133-42

University of Glasgow | Institute of Health & Wellbeing

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

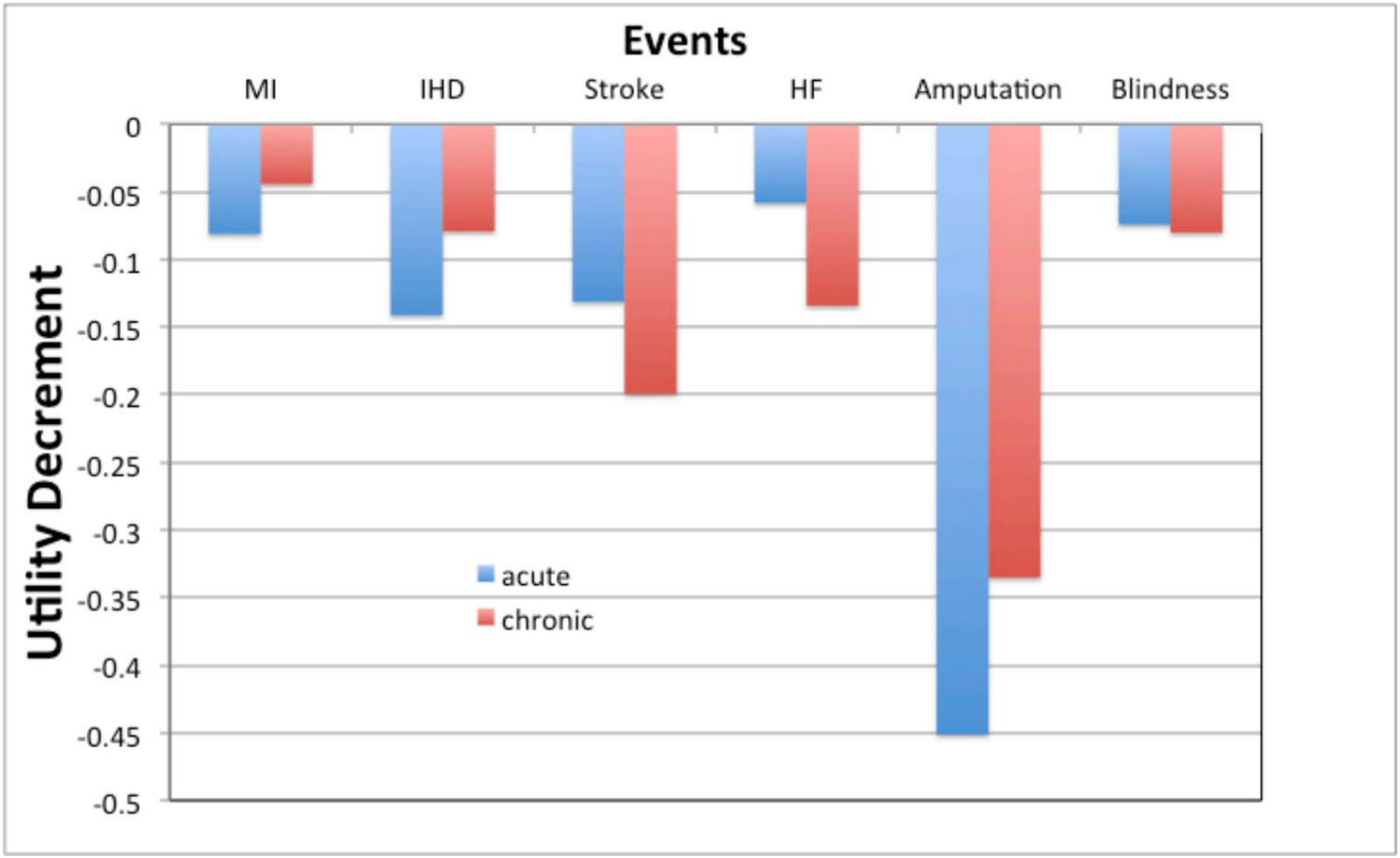
- Analyse utility data by clinical trial arm
- Direct utility or change from baseline
- Test differences in utility between arms

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 / events in the trial?
  - Can a sufficient number of assessments be included, could discontinuation of follow-up introduce bias?
  - 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)

Open discussion

- 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?
- What would you like to see included in an ISPOR Good Research Practice publication?