# Minimal Clinically Important Difference in EQ-5D: We Can Calculate it – But Does That Mean We Should?

May 23, 2017

## ISSUE

- Common to see the calculation of minimally clinically important differences (MCID) in disease specific outcomes, and methods have evolved for doing so.
- As the majority of disease specific outcomes are based on an arbitrary scale, the need to define MCID may be reasonable. But how does this concept translate to a cardinal utility scale such as the EQ-5D?
- Many examples of MCID for EQ-5D now appear in the literature – but do they have meaning?
- Should we even attempt to define MCID for a utility score where the preference weights indicate how much one state is preferred to another?

## OVERVIEW

- In order to be combined with survival estimates, health related quality of life measures need to be anchored at 0 for dead and 1 for full health, and have cardinal utility scale properties.
- Disease-specific health related quality of life measures are not subject to the same constraints, therefore the resulting scales are not comparable between disease areas
- The lack of common scale makes it natural to ask what level of difference on a disease specific scale is “clinically meaningful”– hence the development of methods to determine the MCID.
- However, these methods are increasingly being used to calculate the MCID of utility measures such as EQ-5D; estimates of the EQ-5D MCID now exist across a number of disease areas. But what do these estimates really mean for a generic cardinal utility measure?
- Our panel today represents an important unresolved debate in our field that sits on the intersection between Outcomes Research and PharmacoEconomics.

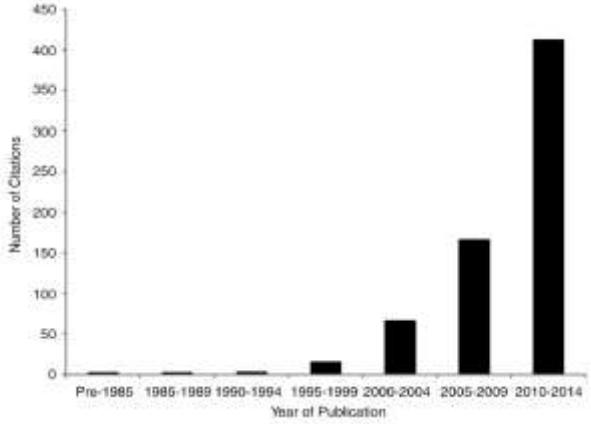


## DEFINITIONS

- The term MCID was first described in 1989.
- “…. The smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management.” (Jaeschke et al.)
- This definition involves two constructs:
  - 1) a minimal amount of patient reported change, and;
  - 2) something significant enough to change patient management.
- MID: minimally important difference
- MCD: minimal clinical difference
- MCSD: minimal clinically significant difference

## GROWING INTEREST IN MCID

Number of citations found in PubMed with search terms of minimal (clinically) important difference, by 5-year stratum.




*Source:* Johnston, B. C., et al. (2015). "Minimally important difference estimates and methods: a protocol." BMJ Open **5**(10): e007953.

5 AVALON HEALTH ECONOMICS

## Ranges of EQ-5D MCID Estimates (Coretti et al. 2014)

| Study | Disease Area | MCID |
|---|---|---|
| Larsen et al. | Musculoskeletal | 0.08 |
| Marra et al. | Musculoskeletal | 0.05 |
| Solberg et al. | Musculoskeletal | 0.30 |
| Soer et al. | Musculoskeletal | 0.03 |
| Parker et al. | Musculoskeletal | 0.24 |
| Parker et al. | Musculoskeletal | 0.14-0.24 |
| Impellizzeri et al. | Musculoskeletal | 0.16 |
| Parker et al. | Musculoskeletal | 0.29-0.52 |
| Parker et al. | Musculoskeletal | 0.15-0.54 |
| McDonough et al. | Musculoskeletal | 0.12-0.15 |
| Boonen et al. | Musculoskeletal | 0.36 |
| Staerkle et al. | Musculoskeletal | 0.36 |
| Kvam et al. | Oncology | 0.08-0.10 |
| Pickard et al. | Oncology | 0.07-0.12 |
| Le et al. | PTSD | 0.04-0.10 |
| Stark et al. | IBD | 0.08-0.11 |
| Shikiar et al. | Psoriasis | 0.09-0.22 |
| Walters & Brazier | Mixed | 0.07 |

Authors found, overall, MCID ranges from 0.03 to 0.54, with a raw average across all 18 studies of 0.18.

6 AVALON HEALTH ECONOMICS

## PANELISTS

- **Moderator:** Cara Scheibling, Associate Director & Partner, Avalon Health Economics, Morristown, New Jersey

- **Panelists:**
  - Andrew Briggs, DPhil, William Lindsay Chair of Health Economics, University of Glasgow, Glasgow, Scotland, UK., Visiting Scholar at Memorial Sloan Kettering, and Director & Partner at Avalon Health Economics;
  - Simon Pickard, PhD, Chair & Executive Committee EuroQol Group, and Professor at University of Illinois, Chicago;
  - Andrew Lloyd, DPhil, Director, AcasterLloyd Consulting Ltd, Oxford, UK

- **Panelist Perspective:**
- Andrew Briggs will argue that MCID should not be translated to QALY calculations or cost-effectiveness
- Simon Pickard will argue that MCID is a relevant concept for PRO and HRQoL
- Andrew Lloyd will present thoughts about the use of MID in rare diseases.


AVALON HEALTH ECONOMICS

# MCID IN EQ-5D: WE CAN CALCULATE IT – BUT DOES THAT MEAN WE SHOULD?

NO!

Andrew Briggs

Memorial Sloan Kettering Cancer Center

University of Glasgow

Under each heading please tick one box that describes your health today

**Mobility**
- I have no problems in walking about ☐
- I have some problems in walking about ☑
- I am confined to bed ☐

**Self-Care**
- I have no problems with self-care ☑
- I have some problems washing or dressing myself ☐
- I am unable to wash or dress myself ☐

**Usual Activities** *(e.g. work, study, housework, family or leisure activities)*
- I have no problems with performing my usual activities ☐
- I have some problems with performing my usual activities ☑
- I am unable to perform my usual activities ☐

**Pain/Discomfort**
- I have no pain or discomfort ☐
- I have moderate pain or discomfort ☑
- I have extreme pain or discomfort ☐

**Anxiety/Depression**
- I am not anxious or depressed ☑
- I am moderately anxious or depressed ☐
- I am extremely anxious or depressed ☐

**MOBILITY**
- I have no problems in walking about ☐
- I have slight problems in walking about ☑
- I have moderate problems in walking about ☐
- I have severe problems in walking about ☐
- I am unable to walk about ☐

**SELF-CARE**
- I have no problems washing or dressing myself ☑
- I have slight problems washing or dressing myself ☐
- I have moderate problems washing or dressing myself ☐
- I have severe problems washing or dressing myself ☐
- I am unable to wash or dress myself ☐

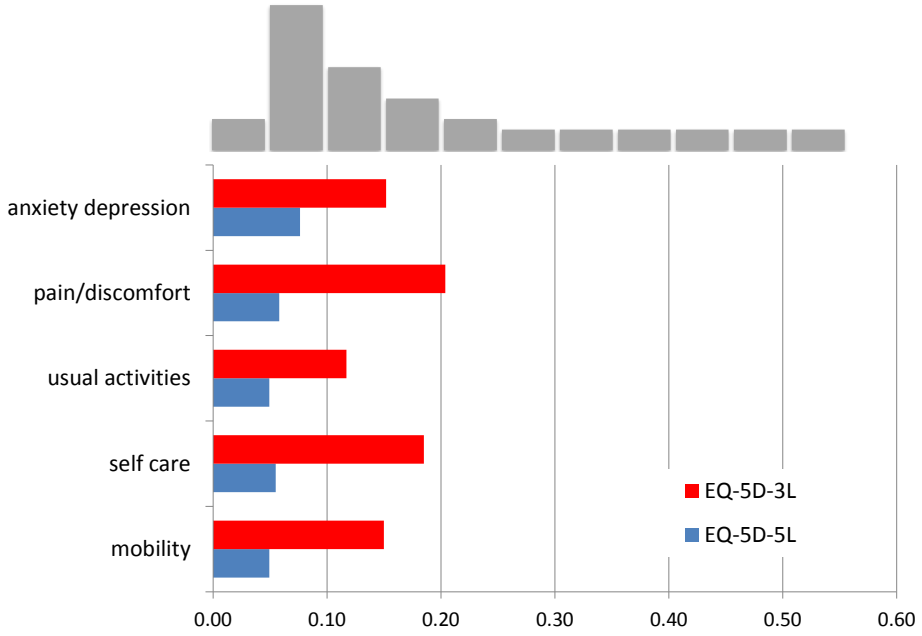
**USUAL ACTIVITIES** *(e.g. work, study, housework, family or leisure activities)*
- I have no problems doing my usual activities ☐
- I have slight problems doing my usual activities ☐
- I have moderate problems doing my usual activities ☑
- I have severe problems doing my usual activities ☐
- I am unable to do my usual activities ☐

**PAIN / DISCOMFORT**
- I have no pain or discomfort ☐
- I have slight pain or discomfort ☑
- I have moderate pain or discomfort ☐
- I have severe pain or discomfort ☐
- I have extreme pain or discomfort ☐

**ANXIETY / DEPRESSION**
- I am not anxious or depressed ☐
- I am slightly anxious or depressed ☑
- I am moderately anxious or depressed ☐
- I am severely anxious or depressed ☐
- I am extremely anxious or depressed ☐



# EQ-5D level changes from full health



## Using Instrument-Defined Health State Transitions to Estimate Minimally Important Differences for Four Preference-Based Health-Related Quality of Life Instruments

Luo, Nan PhD*; Johnson, Jeffrey A. PhD†; Coons, Stephen Joel PhD‡

Medical Care: April 2010 - Volume 48 - Issue 4 - pp 365-371
doi: 10.1097/MLR.0b013e3181c162a2
Original Article

Abstract | Author Information

Objective: To estimate minimally important differences (MIDs) for the EQ-5D, Health Utilities Index Mark II (HUI2), HUI3, and SF-6D health index scores using health-state transitions defined by each instrument's multiattribute health classification (MAHC) system.

Methods: We assume that changes in preference scores associated with the smallest health transitions defined by an MAHC system are minimally important. Any transitions between 2 health states defined by an MAHC system which differ in only one health dimension or attribute and by only one functional level are considered "smallest health transitions." Thus, each such health transition provides 1 MID estimate. The MID for each of the 4 instruments was estimated using all the hypothetical smallest health transitions defined by its MAHC system.

Results: Based on our definitions, the total number of smallest health transitions was 405 for the EQ-5D, 127,600 for the HUI2, 6,382,800 for the HUI3, and 86,700 for the SF-6D. The mean (standard deviation) MID estimate was 0.040 (0.026) for the EQ-5D (US algorithm), 0.082 (0.032) for the EQ-5D (UK algorithm), 0.045 (0.039) for the HUI2, 0.032 (0.027) for the HUI3, and 0.027 (0.028) for the SF-6D. The effect sizes of these MID estimates ranged from 0.11 to 0.37. These MID estimates are quite comparable to published values estimated from empirical data using anchor-based methods.

Conclusions: It is possible to use health transitions defined by the MAHC system to estimate the MIDs for preference-based health index scores. This study provides new information regarding MID estimates for the 4 health indices examined.

## Using Instrument-Defined Health State Transitions to Estimate Minimally Important Differences for Four Preference-Based Health-Related Quality of Life Instruments

Luo, Nan PhD*; Johnson, Jeffrey A. PhD†; Coons, Stephen Joel PhD‡

Medical Care: April 2010 - Volume 48 - Issue 4 - pp 365-371
doi: 10.1097/MLR.0b013e3181c162a2
Original Article

Abstract | Author Information

Objective: To estimate minimally important differences (MIDs) for the EQ-5D, Health Utilities Index Mark II (HUI2), HUI3, and SF-6D health index scores using health-state transitions defined by each instrument's multiattribute health classification (MAHC) system.

Methods: We assume that changes in preference scores associated with the smallest health transitions defined by an MAHC system are minimally important. Any transitions between 2 health states defined by an MAHC system which differ in only one health dimension or attribute and by only one functional level are considered "smallest health transitions." Thus, each such health transition provides 1 MID estimate. The

Results: Based on our definitions, the total number of smallest health transitions was 405 for the EQ-5D, 127,600 for the HUI2, 6,382,800 for the HUI3, and 86,700 for the SF-6D. The mean (standard deviation) MID estimate was 0.040 (0.026) for the EQ-5D (US algorithm), 0.082 (0.032) for the EQ-5D (UK algorithm), 0.045 (0.039) for the HUI2, 0.032 (0.027) for the HUI3, and 0.027 (0.028) for the SF-6D. The effect sizes of these MID estimates ranged from 0.11 to 0.37. These MID estimates are quite comparable to published values estimated from empirical data using anchor-based methods.

Conclusions: It is possible to use health transitions defined by the MAHC system to estimate the MIDs for preference-based health index scores. This study provides new information regarding MID estimates for the 4 health indices examined.

**Using Instrument-Defined Health State Transitions to Estimate Minimally Important Differences for Four Preference-Based Health-Related Quality of Life Instruments**

Luo, Nan PhD*; Johnson, Jeffrey A. PhD†; Coons, Stephen Joel PhD‡

Medical Care: April 2010 - Volume 48 - Issue 4 - pp 365-371
doi: 10.1097/MLR.0b013e3181c162a2
Original Article

Abstract | Author Information

Objective: To estimate minimally important differences (MIDs) for the EQ-5D, Health Utilities Index Mark II (HUI2), HUI3, and SF-6D health index scores using health-state transitions defined by each instrument's multiattribute health classification (MAHC) system.

Methods: We assume that changes in preference scores associated with the smallest health transitions defined by an MAHC system are minimally important. Any transitions between 2 health states defined by an MAHC system which differ in only one health dimension or attribute and by only one functional level are considered [illegible] health transition provides 1 MID estimate. The

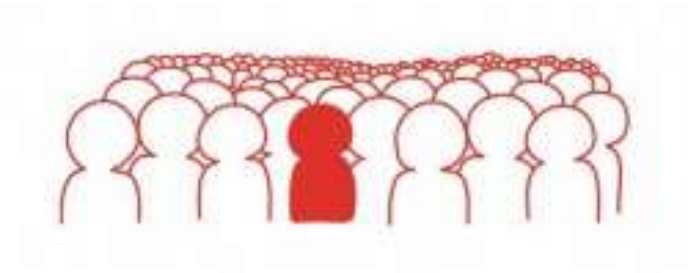
Results: Based on [illegible] MID estimate was [illegible] smallest health transitions was 405 for the EQ-5D, 127,600 for the HU[illegible]00 for the SF-6D. The mean (standard deviation) MID estimate was [illegible] 0.045 (0.039) for th[illegible] these MID estimate[illegible] values estimated from empirical data using anchor-based methods.

MID estimate was

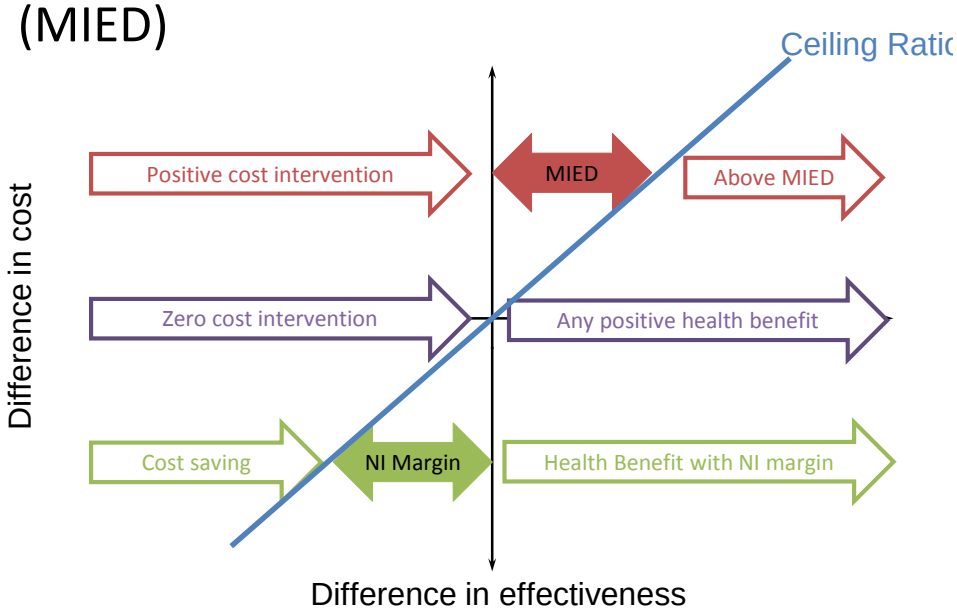
0.082 (0.032) for the EQ-5D (UK algorithm),

Conclusions: It is possible to use health transitions defined by the MAHC system to estimate the MIDs for preference-based health index scores. This study provides new information regarding MID estimates for the 4 health indices examined.

All it takes is one person…



# Minimally Important Economic Difference (MIED)




MINIMAL CLINICALLY IMPORTANT DIFFERENCE IN EQ-5D-5L: WE CAN CALCULATE IT, BUT DOES THAT MEAN WE SHOULD?

YES!

A. Simon Pickard, PhD

*Professor, University of Illinois at Chicago*

# Acknowledgements

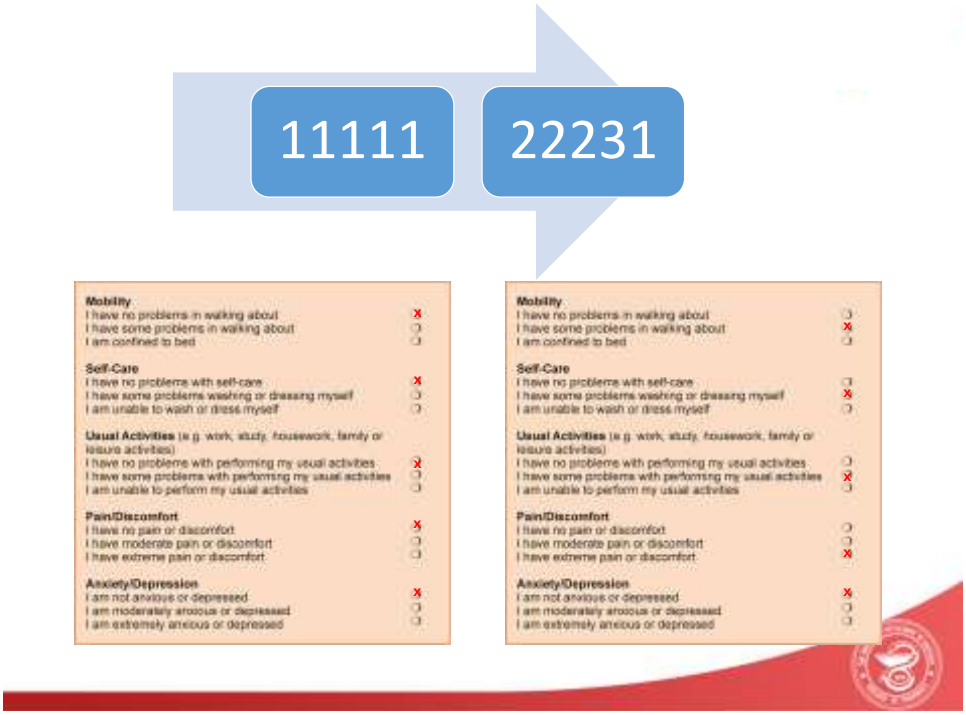

- Michael Herdman
- Jeffrey A Johnson
- David Whitehurst
- Kim Rand-Hendriksen


TRIALS

COMMENTARY Open Access

Trial-based clinical and economic analyses: the unhelpful quest for conformity

- Clinical and economic evaluation do not need to be compatible
- For EEs, both cost and outcome are jointly considered




# What is a meaningful difference?

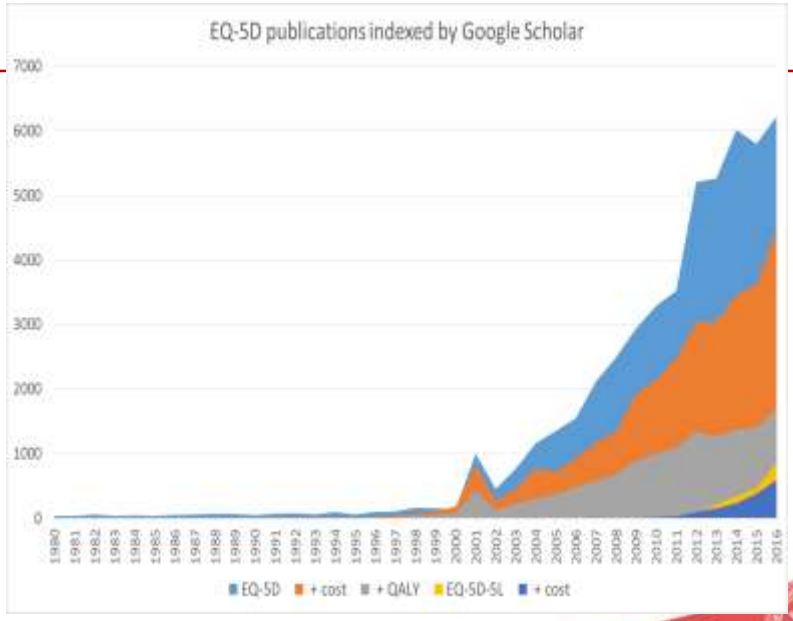
- "the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management"
  - Jaeschke R, Singer J, Guyatt GH. Control Clin Trials 1989;10:407-415.
- Or make you contemplate a visit to the doctor..

## When might change in components of EQ-5D be important?

- Descriptive system
  - Movement on any level ("health state transition")
- Value sets
  - On individual level, any change in score if based on descriptive system with weights for:
    - General population (societal) weights
    - Patient weights
    - Any other sub-group of interest
- VAS


## Measuring and Valuing Health

- Preference-based measures of health are important to HTA, e.g.
  - EQ-5D
  - Health Utilities Index
  - SF-6D
- Societal preference-weights ("value sets") facilitate QALY calculations in cost-utility analysis -> inform resource allocation
- However, they have many other applications




Slide courtesy of Kim Rand-Hendriksen

## Non-economic applications of EQ-5D

- Stand-alone HRQoL measure
  - Burden of illness
  - Cohort studies
  - Clinical trials
- Population health surveys
  - Population monitoring
  - Comparative indicator (between countries, between groups, evidence of inequities)
- Routine Outcome Measurement (ROM)
  - Meso-level: indicator of quality of care / evaluate outcomes of care
  - Patient-level: individual monitoring
- EQ-5D reported as an “extra”

## Interpretation and Knowledge Translation

- User guidance and support:
  - How do I score the measure?
  - How do I interpret the measure?
  - What delta should I use when planning my study?
  - What decisions are being made based on this score/metric?

## Assessing HRQL instruments: attributes and review criteria

1) Conceptual & Measurement Model
2) Reliability
3) Validity
4) Responsiveness
5) Interpretability
6) Burden
7) Alternative Forms
8) Cultural and Language Adaptations

Scientific advisory committee of MOT. Qual Life Res 2002; 11:193-205.

## Statistical Significance vs Clinical Importance

| | | Statistically Significant (e.g. p<0.05) | |
|---|---|---|---|
| | | **Yes** | **No** |
| **Clinically Important (magnitude)** | **Yes** | Good, Something worthwhile | May still be an important outcome (power?) |
| | **No** | Trivial | No good, not something to invest in |

Slide courtesy of Jeff Johnson

## Approaches to MIDs

- **Anchor-based approach:** use changes in health status measures and *a priori* defined criteria to identify *small/minimally important change group*
- **Distribution-based:** related to SD/Effect Size
- **Instrument-defined approach:** use EQ-5D-5L health state and scoring algorithm to quantify difference in index score between baseline health state and *single-level transitions*

# EQ-5D-3L UK scoring algorithm

| Full health | 1.000 |
|---|---|
| Constant term (for any dysfunction state) | -0.081 |
| Mobility level 2 | -0.069 |
| Mobility level 3 | -0.314 |
| Self-care level 2 | -0.104 |
| Self-care level 3 | -0.214 |
| Usual activities level 2 | -0.036 |
| Usual activities level 3 | -0.094 |
| Pain/discomfort level 2 | -0.123 |
| Pain/discomfort level 3 | -0.386 |
| Anxiety/depression level 2 | -0.071 |
| Anxiety/depression level 3 | -0.236 |
| N3 (level 3 occurs for at least one dimension) | -0.269 |

29

# Instrument Defined (Health Transition Approach)

- Luo et al (2010): first published instance of the use of health state transitions (HST) to estimate MIDs.
- Assumes that changes in preference scores associated with the smallest health transitions defined by an MAHC system are minimally important,
    - i.e. transitions between 2 health states which differ in only one health dimension or attribute and by only one functional level are considered “smallest health transitions.”
- For EQ-5D, excluded levels 2 to 3 transitions because they represent a substantial change (e.g. ‘some problems walking’ to ‘confined to bed’).
- The mean (SD) MID estimate was 0.040 (0.026) for US algorithm and 0.082 (0.032) for the UK

# FDA PRO guidance

**Guidance for Industry**

**Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims**

3. Planning for Clinical Trial Interpretation Using a Responder Definition

Regardless of whether the primary endpoint for the clinical trial is based on individual responses to treatment or the group response, it is usually useful to display individual responses, often using an *a priori* responder definition (i.e., the individual patient PRO score change over a predetermined time period that should be interpreted as a treatment benefit). The responder definition is determined empirically and may vary by target population or other clinical trial



*Concepts Needing Recommendations*

design characteristics). Therefore, we will evaluate an instrument's responder definition in the context of each specific clinical trial.

The empiric evidence for any responder definition is derived using anchor-based methods. Anchor-based methods explore the associations between the targeted concept of the PRO instrument and the concept measured by the anchors. To be useful, the anchors chosen should be easier to interpret than the PRO measure itself. For example, the number of exacerbation episodes collected as an exacerbation diary has been used to determine a responder definition for PRO instruments assessing the *time/course of exacerbations*. A 10 percent reduction in exacerbation episodes might be proposed as the anchor for defining a responder on the PRO instrument. Confirmation of this anchor approach in early clinical trials can provide the basis for the proposed responder definition in the confirmatory trials.

Another anchor-based approach in defining responders makes use of patient ratings of change (enhancement or different periods of time or upon exit from a clinical trial). These numerical ratings range from *worse* to the *same* and *better*. The difference in the PRO score for persons who rate their condition the same and better or worse can be used to define responders to treatment. Patient ratings of change are less useful as anchors when patients are not blinded to treatment assignment.

Another set of approaches to defining a responder are distribution-based methods that use, for example, the between-person standard deviation or the standard error of measurement to define a meaningful change on a scale. Distribution-based methods can be used to categorize these changes as small, moderate, and large and often can be combined with anchor-based estimates to provide confidence in the responder definition. Distribution-based methods for determining clinical significance of particular score changes should be considered as supportive and are not appropriate as the sole basis for determining a responder definition.

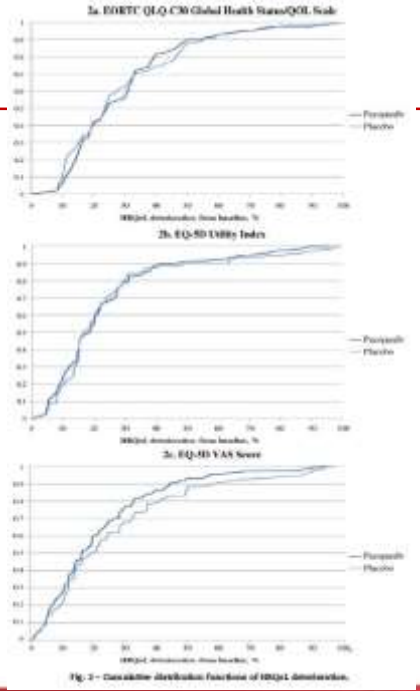



Fig. 2 – Cumulative distribution functions of HRQoL deterioration.

Cella D, Pickard AS, et al., Health-related quality of life in patients with advanced renal cell carcinoma receiving pazopanib or placebo in a randomised phase III trial, Eur J Cancer (2011), doi:10.1016/j.ejca.2011.05.017

## Summary

- There is a lot of subjectivity in interpretation
- No ideal approach to MIDs
- But without any guidance for interpretation….



# Is there a role for establishing an important change on EQ-5D?

Andrew Lloyd

Acaster Lloyd Consulting Ltd

## Cost effectiveness

- EQ-5D is a useful way of measuring health/ HRQL
- Allows for standardisation in submissions
- EQ-5D used in
  - Economic evaluation
  - Routine Outcome Measurement (PROMS)
  - Clinical trials
- Heard clear arguments that MID estimates
  - No application within cost effectiveness analysis
  - Potentially useful for interpretation in clinical applications

## Could estimates of important change support economic evaluation?

- May be special cases where an estimate of important change could support an economic evaluation
- One case is in rare diseases
- NICE and others are working on a large range of orphan drug reviews

# Orphan drugs & HTA process

- A lot of factors make the assessment of orphan drugs particularly challenging
- Trial designs
  - Very small, often single arm, heterogeneity in HRQL
- Cost effectiveness
  - Drug costs often very high; but treatment often conveys huge health gains
  - Not close to standard criteria of cost effectiveness
- Value
  - Many orphan drugs are the only treatment available in a condition
  - Large unmet need
  - Huge potential also for opportunity cost
- Scale of this problem likely to grow

# Reimbursement decisions

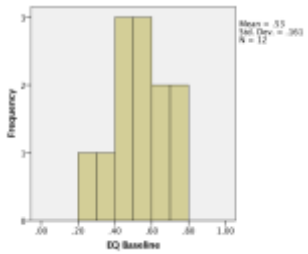
- Health systems cannot afford to approve access for all orphan drugs
  - 10% of US drug spend on orphan drugs
- NICE et al are left to make a decision
  - Assess overall health benefit
  - Cost effectiveness
  - Overall budget impact
- Health benefit assessed by QALYs
  - How much health do we get for our money?
- Despite limitations in data a decision is still needed
  - Estimating important change may help inform that decision

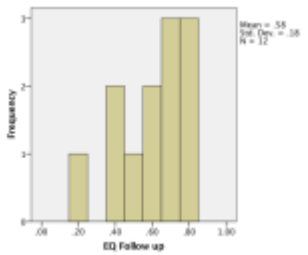
# Assessment of utilities in rare disease

- Utility (EQ-5D) data often extremely limited
  - Aggregating data from just a few people
  - No comparison data
  - Are data representative?
  - Can we infer?

# A Case study – PDQ1 inhibitor

- Data from 12 patients (no controls)
- Are we confident making inferences from these data?
  - Heterogeneous
  - Mean change small
  - Considerable uncertainty
- Adding information may reduce decision uncertainty



# Use of responder definition

- Estimate a degree of change on EQ-5D that is important for an individual
  - Response
- Responder definition = qualitative change in a patient
- Could be change moving from
  - *Some problems walking about* to *No problems walking about*
- Classify patients according to definition of response

# PDQ-1 case study

- Here a responder definition used
- Using this approach
  - 9 responded to therapy
  - 3 showed no response
- Provides an alternative interpretation of results to support decision

| Baseline | Follow up | |
|---|---|---|
| 0.55 | 0.65 | **R** |
| 0.40 | 0.20 | **NR** |
| 0.65 | 0.40 | **NR** |
| 0.70 | 0.80 | **R** |
| 0.55 | 0.65 | **R** |
| 0.55 | 0.75 | **R** |
| 0.42 | 0.55 | **R** |
| 0.69 | 0.80 | **R** |
| 0.78 | 0.80 | **NR** |
| 0.25 | 0.40 | **R** |
| 0.34 | 0.50 | **R** |
| 0.44 | 0.60 | **R** |

Responder definition = 0.10

# Use of responder definitions

- Decision makers often faced with sub-optimal datasets
- Assessing health gain of orphan drugs can be very challenging
- Applying a responder definition can provide alternative way to interpret data
  - Doesn't require additional data to be collected
  - May support decision making

# Questions?