ISPOR Asia Pacific 2018 Keio Plaza Hotel, Tokyo, Japan
Sunday, 9 September 2018, 3:45 PM–4:45 PM @ Room: Nishiki

EQ-5D

W2: THE EQ-5D-5L INSTRUMENT: PAST, PRESENT AND FUTURE

Discussion Leaders:

Mark Oppe
EuroQol Research Foundation, Rotterdam, The Netherlands
Nan Luo
National University of Singapore, Singapore
Ataru Igarashi
The University of Tokyo, Japan
Kim Rand
University of Oslo, Norway

EuroQol Group

EQ-5D

Conflict of interest & disclaimer

- The discussion leaders are members of the EuroQol Group, a not-for-profit international research organization
- The views of the discussion leaders expressed in the workshop do not necessarily reflect the views of the EuroQol Group

The past & present: how was EQ-5D-5L developed?

EuroQol Group

Mark Oppe

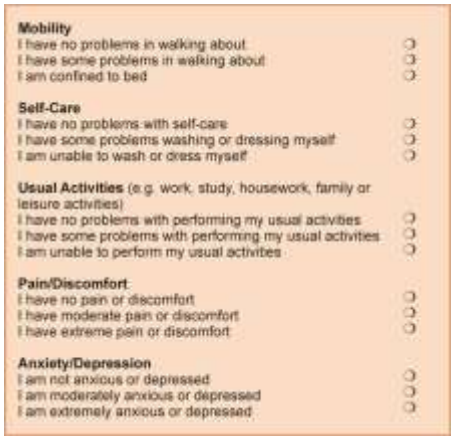
oppe@euroqol.org

Overview

- Background
- Changes in the EQ-5D descriptive system
- Changes in the EQ-5D valuation system
- Conclusions

Why move from 3 to 5 levels?

- **EQ-5D-3L robust and well-validated, but…**
  - Ceiling effects in some populations
  - Perceived as insensitive to small changes in health status or between-group differences, particularly in milder disease
  - Literature and previous research with experimental expanded versions suggested advantages to having more severity levels
- **Decisions made**
  - Retain the same dimensions
  - Expand to 5 levels of severity



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

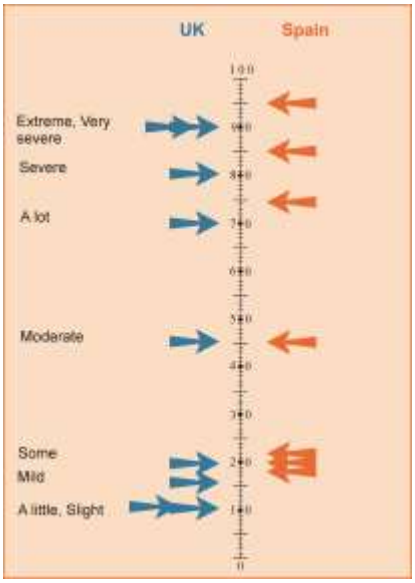
Anxiety/Depression
I am not anxious or depressed
I am moderately anxious or depressed
I am extremely anxious or depressed

Developing the EQ-5D-5L descriptive system

Two-stage approach

- Response scaling
  - In UK and Spain to select severity labels for 5L version
- Focus groups
  - To assess face and content validity of new versions in each country

Response scaling



EQ-5D-5L

- Top level for mobility changed:

3L: “Confined to bed”

5L: “Unable to walk about”

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

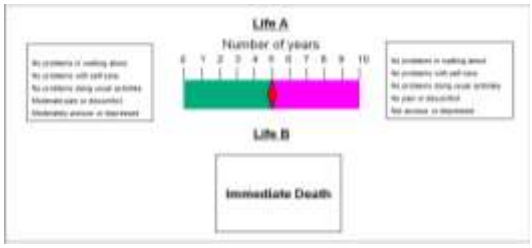
USUAL ACTIVITIES (e.g. work, study, housework, family or leisure activities)
I have no problems doing my usual activities
I have slight problems doing my usual activities
I have moderate problems doing my usual activities
I have severe problems doing my usual activities
I am unable to do my usual activities

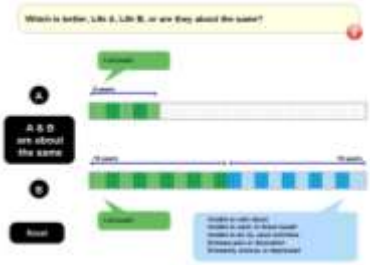
PAIN / DISCOMFORT
I have no pain or discomfort
I have slight pain or discomfort
I have moderate pain or discomfort
I have severe pain or discomfort
I have extreme pain or discomfort

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

Valuation

- Valuation tasks
  - 3L: ranking + VAS as warm up for TTO
  - 5L: wheelchair example + 3 practice states as warm up for TTO
  - 5L: DCE task
- TTO framework
  - 3L: conventional TTO
  - 5L: composite TTO





Experimental design

- **3L VAS & TTO**
  - 43 states
  - Blocks of 13 states
  - Manually selected
- **5L TTO**
  - 86 states
  - Blocks of 10 states
  - Optimisation algorithm
- **5L DCE**
  - 196 pairs
  - blocks of 7 pairs
  - Optimisation algorithm

Data collection

- **3L**
  - F2F personal interviews
  - TTO board, paper VAS
  - Pen and paper data recording

- **5L**
  - Computer assisted F2F personal interviews
  - standardised training sessions of PIs
  - standardised interviewer training materials
  - standardised interviewer script

Quality control

- **3L:** QC not formalised in protocols
- **5L:** Cyclic quality control process during entire data collection period
  1. protocol compliance
  2. interviewer effects
  3. assessment of overall data quality
  4. Interim analysis at n=250, 500, 750

Exclusion of data

- **3L**
  - at the respondent level, based on interviewer judgement
  - at the respondent level, based on study team judgment of characteristics of data
- **5L**
  - at interviewer level (consistently failed to pass QC)
  - at respondent level, based on interviewer judgement
  - at observation level, based on respondent judgement (feedback module)

Modelling

3L

- Rescaling of WTD values
- OLS or random effects models for TTO
- Main effects + "interaction" terms (e.g. N3)
- Bounded nature of distribution not taken into account
- Heteroscedastic nature of data not taken into account

5L

- TTO only or TTO+DCE Hybrid models
- Main effects only
- Bounded nature of distributions taken into account
- Heteroscedastic nature of data taken into account

Summary

- Descriptive system:
  - 2 additional levels
  - Changed top level label for mobility
- Valuation:
  - Tasks
  - Experimental designs
  - Quality control
  - Modeling

The present: how does EQ-5D-5L perform?

EuroQol Group

Nan Luo

ephln@nus.edu.sg

Does 5L have better measurement properties than 3L?

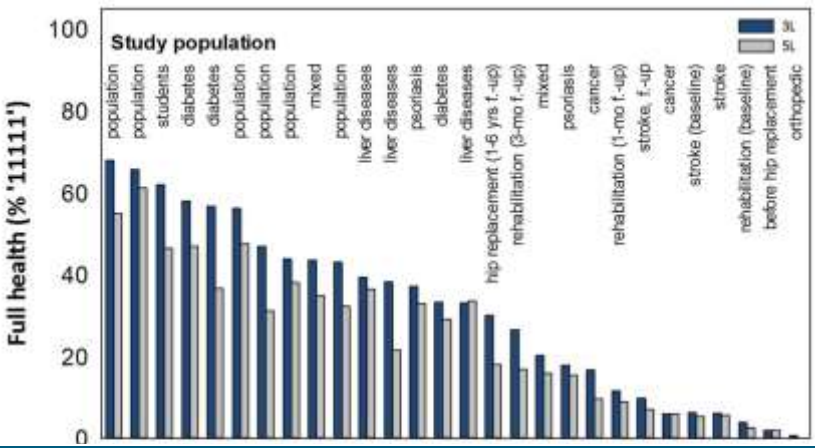
The 5 dimensions

- Ceiling effects
- Distributions
- Reliability
- Sensitivity to change

The utility index

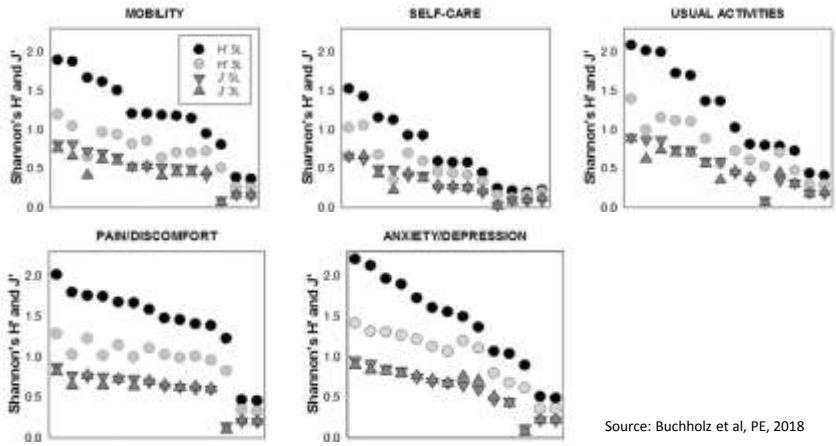
- Distributions
- Reliability
- Sensitivity to difference
- Sensitivity to change

Ceiling effects of 5L and 3L in comparative studies (n=25)



Source: Buchholz et al, PE, 2018

Shannon's H' and J' of 3L and 5L dimensions in comparative studies (N=14)



Source: Buchholz et al, PE, 2018

Test-retest reliability of 5L and 3L in comparative studies (N=4)

Sample (size)	Mean time interval	Value set	Reliability of dimensions (*Kappa*)	Reliability of index score (ICC)
South Korean cancer patients (n = 78)	11.5 days (IQR 6–15)	3L: South Korea 5L: mapping	3L = 0.39–0.66 5L = 0.36–0.64	3L = 0.75 5L = 0.77
South Koreans from the general population (n = 100)	18.7 days (SD 4.5)	3L: South Korea 5L: mapping	3L = 0.31–0.64 5L = 0.33–0.69	3L = 0.61 5L = 0.75
Chinese hepatitis B patients (n = 120)	1 week	3L: Japan 5L: mapping	3L = 0.74–0.93 5L = 0.73–0.98	3L = 0.83 5L = 0.93
Thai diabetes patients treated with insulin (n = 117)	14–21 days	3L: Thailand 5L: mapping	3L = 0.39–0.70 5L = 0.44–0.57	3L = 0.64 5L = 0.70

Source: Buchholz et al, PE, 2018

Sensitivity to difference of 5L versus 3L index score in comparative studies (N=6)

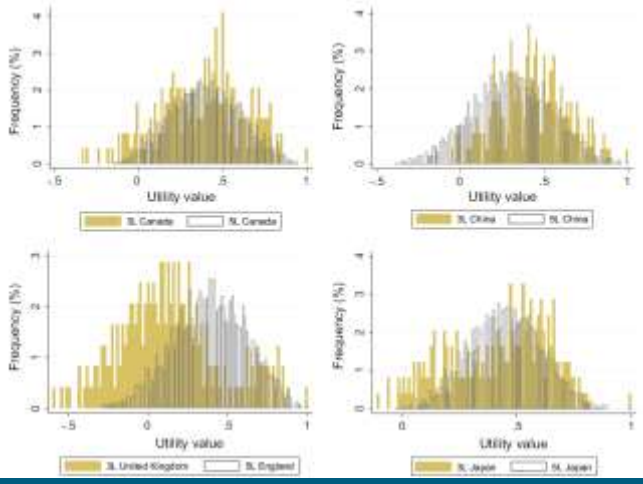
Study population	Definition of known-groups (relative efficiency of 5L vs 3L)	No. of comparisons	Sensitivity to difference
General Greek population over 40 years old	Age (1.41), BMI (1.90), chronic condition (1.34), physician visit (1.20)	4	5L > 3L
General Portugal population aged 30 years or below	Chronic condition (0.997)	1	3L > 5L
Greek psoriatic patients	Severity (1.04), age (1.03), BMI (1.40), 5 comorbidities (1.00-1.39)	8	5L > 3L
Singaporean diabetic patients	Comorbidity (1.92), complication (0.73), BMI (2.13), HbA1c (4.00)	4	5L > 3L
Chinese diabetic patients	7 compilations/comorbidities (1.08-6.10), cardiovascular diseases (0.85), glucose (1.55), BMI (1.13)	10	5L > 3L
1 European student group and 8 patient groups	Chronic condition	8	3L > 5L
	Disease burden	12	5L > 3L

Sensitivity to change of 5L and 3L in comparative studies (N=5)

Sample (size)	Effect measure	Time interval	Value set	Sensitivity of index score	Sensitivity of dimensions
Chinese hepatitis B patients (n = 120)	Laboratory and blood test defined improvement	1 week	3L: Japan 5L: mapping	3L > 5L	
German rehabilitation patients (n_{t1-2} = 224 and n_{t1-3} = 154)	Improvement by rehabilitation	t_1: Beginning, t_2: End of, t_3: 3 months after rehabilitation			5L > 3L
Polish stroke patients (n = 112)	mRS- and BI-defined deterioration and improvement	t_1: 1 week t_2: 4 months poststroke	3L: Poland 5L: mapping	3L > 5L	
Singaporean cataract patients (n = 148)	Improvement by cataract surgery and confirmed by visual acuity	t_1: Before surgery t_2: 3 months after surgery	3L: UK, Singapore 5L: UK; mapping	5L > 3L	
Canadian patients receiving total hip/knee replacement surgery (3L: n = 1,594; 5L: n = 3,180)	WOMAC defined deterioration and improvement	t_1: Before surgery t_2: 3 months after surgery	3L: Canada 5L: Canada	5L > 3L	

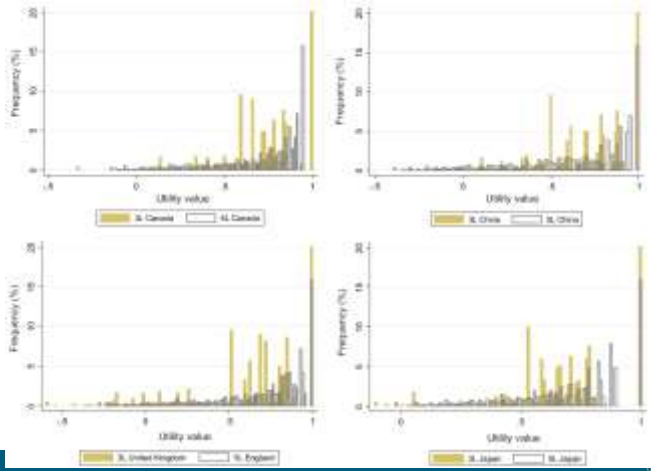
Source: Buchholz et al, PE, 2018

Histogram of all possible 5L (N=3,125) and 3L (N=243) values in 4 countries



Source: Janssen et al, PE, 2018

Histograms of utilities of a study cohort (N=3,467) using value sets of 4 countries



Source: Janssen et al, PE, 2018

Impact on economic evaluation

- EQ-5D data is widely used to calculate quality-adjusted life years (QALYs) in economic evaluation of new health technologies
- Better sensitivity in measuring health outcomes means more precise quantification of QALYs, which may be higher or lower than imprecise QALYs estimates.
- Therefore, it is desirable to ascertain the impact of switching from EQ-5D-3L to EQ-5D-5D on economic evaluation results, even though the latter is more sensitive than the former

Case studies comparing the effect of 5L and 3L values on cost-utility analysis

- Using 3L data to crosswalk to 5L values, Hernandez et al (2018) found that moving from 3L to 5L caused a decrease of up to 87% in incremental QALYs gained in almost all CUA cases from the UK.
- Using 5L data and a CUA of dialysis modalities for end-stage renal disease, Yang et al (forthcoming) found that the impact of switching form 3L to 5L on QALYs gained and ICER depends on many factors including the value sets used.

Summary

- EQ-5D-5L appears to have better measurement properties than EQ-5D-3L.
- The switch from EQ-5D-3L and EQ-5D-5L could impact on the economic evaluation results, posing a challenge to policy makers and researchers who have been using EQ-5D-3L to conduct cost-effectiveness analysis.

The use of EQ-5D-5L in Asian HTA systems

EuroQol Group

Ataru Igarashi

atarui1@mac.com

Use of EQ-5D questionnaire in the HTAs in Asia-Pacific

- Several HTA guideline have the “Preferred” questionnaire for the data source of utility values
- In general, ”Domestic” utility values are more preferred, while it is not mandatory (Japan, Korea, Australia, Thailand, Malaysia)

Component	“Nationality”
Clinical Efficacy	Higher evidence level would be better, regardless of nationality
Utility value	Domestic data preferred, unless no data is available
Costs	MANDATORY to be domestic

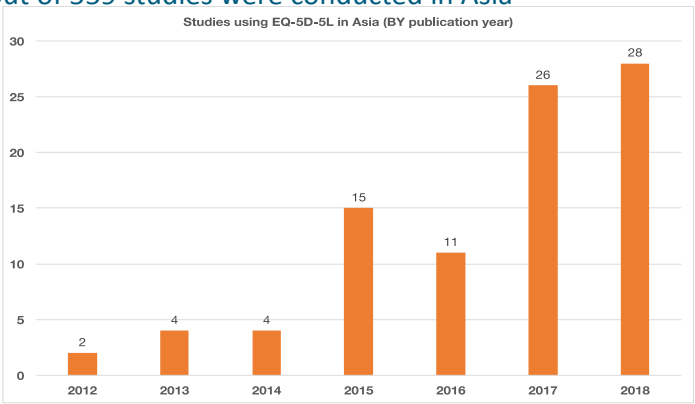
What is the position of the EQ-5D?

- TWO types of prioritization are available when choosing questionnaire among MAUIs (Multi attribute utility index)

IMPLICIT Recommendation	Prefer questionnaire with domestic tariff	Japan, Korea, Australia
EXPLICIT Recommendation	Explicitly mentioning EQ-5D	Malaysia, Thailand

Studies around EQ-5D-5L in Asia (year)

- 90 out of 559 studies were conducted in Asia



Studies around EQ-5D-5L in Asia (by countries/areas)

Country/Year	2012	2013	2014	2015	2016	2017	2018	Total
Brunei					1			**1**
China		1	2	3	2	4	4	**16**
Egypt							1	**1**
Hongkong	1					4	3	**8**
Indonesia						1	3	**4**
Iran							1	**1**
Japan				3	3	3	3	**12**
Korea	1	1	1		1	5	1	**10**
Malaysia							1	**1**
Palestine				2	1	1	1	**5**
Singapore		2	1	5	1	5	3	**17**
Taiwan						1		**1**
Thailand				2	1	1	3	**7**
Vietnam					1	1	4	**6**
Total	**2**	**4**	**4**	**15**	**11**	**26**	**28**	**90**

Studies around EQ-5D-5L in Asia (by types)

Types	2012	2013	2014	2015	2016	2017	2018	**total**
health state	2		3	11	7	14	17	**54**
mapping			1	2		4	5	**12**
methodology		4		2	2	5	1	**14**
valuation					2	3	5	**10**
total	**2**	**4**	**4**	**15**	**11**	**26**	**28**	**90**

Pitfall around searhching Asian QOL survey

To potential “upgrade” for questionnaire

- What kind of dimension(s) can be added to the EQ-5D-5L? (N=24)

Having a good sleep	9
Can take food routinely (3 times per day) Can enjoy every time we have meals	7
Do not have fever	1
Do not taking any medication	1
Good in mental health, not in physical health	1
NONE	5

Various subversion of EQ-5D (-5L)

- Proxy, Youth, several digital version

Proxy-1	**Respondents evaluate patient’s QOL from Respondents’ own view**
Proxy-2	**Respondent evaluate patient’s QOL thinking about “What if patients themselves can fulfill the questionnaire”**
Youth	**Used for children (8-15 years of age)**

EQ-5D

Complicated characteristic system in Japan…

- 3 Types of characters are used (Hiragana/Katakana only used in Japan, while Kanji is done in China or Korea)
- Superscript (RUBY) was inserted for EQ-5D-Y-JP

次の 5 つの質問について、今日のあなたの体やこころの調子にいちばん近いもの 1 つに○をつけてください。

歩きまわるのはどのくらいたいへんですか。
( ) 歩きまわるのはたいへんではない
( ) 歩きまわるのは少したいへんだ
( ) 歩きまわるのはとてもたいへんだ

自分で体をあらったり、着がえをするのはどのくらいたいへんですか。
( ) 自分で体をあらったり、着がえをするのはたいへんではない
( ) 自分で体をあらったり、着がえをするのは少したいへんだ
( ) 自分で体をあらったり、着がえをするのはとてもたいへんだ

EQ-5D

Two EQ-5D versions: challenges and solutions

Kim Rand

kim.rand@ahus.no

EuroQol Group

Questions from users

- Which version should I use?
- We are conducting a follow-up of a study in which we used the -3L. Should we switch to the -5L?
- There is no -5L value set available for [country] (yet). Should we continue using the -3L?

Challenges for regulators

- Different instrument versions mean
  - Different values
  - Different estimates of QALYs gained
  - Different cost/QALY-estimates
  - Potentially different recommendations for priority setting
  - Allowing both versions may result in gaming

Which version is better?

- Quick answer: the -5L appears to be an improvement over the -3L
  - More sensitive to smaller changes in health
  - Fewer respondents in state 11111
  - Better valuation studies, improved quality control
- However:
  - Moving from -3L to -5L may result in fewer QALYs gained (particularly in the UK)
  - There are cases where -3L displays greater sensitivity
  - Changing version and/or value set may impair comparability and may alter funding decisions

Two sets of issues

- Issues related to changing value sets

  Primary concern of the industry and regulators

- Issues related to changing the descriptive system

  Researchers may be more concerned with alterations to the descriptive system.

Changing the descriptive system

- Main concern: loss of backward comparability
- We cannot observe from responses on the -5L which individuals indicating "slight problems" now would have indicated "moderate problems" using the -3L
- The reverse is also true: "no problems" or "moderate problems" on the -3L could both have been "slight problems" on the -5L.
- Approximate proportions may be estimated from studies including both versions

"Which instrument should I use?"

Issue: New study. Concern: existing literature uses the -3L

- Recommendation:
  - Use the -5L
  - Your study will be cited by future -5L studies.
  - Using the -3L propagates the problem to the next study in line.
  - Compare findings with existing literature using crosswalk value sets.

“Which instrument should I use?”

Issue: Longitudinal study with previous -3L data.

- Recommendation:
  - If comparison is crucial and the longitudinal observation is nearly complete, -3L may be considered.
  - If most of the study is in the future, go for the -5L.
  - May consider including both versions temporarily to provide a substantial link.

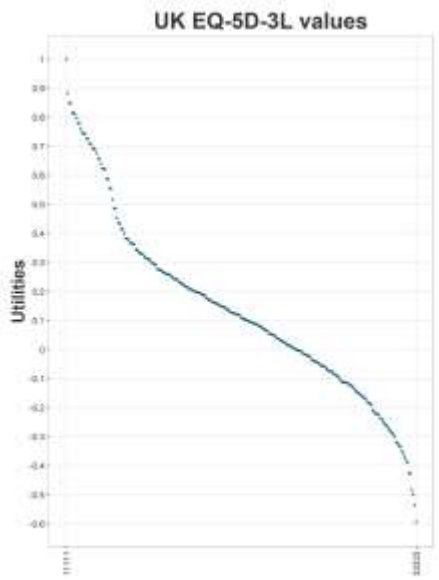
“Which instrument should I use?” cont’d

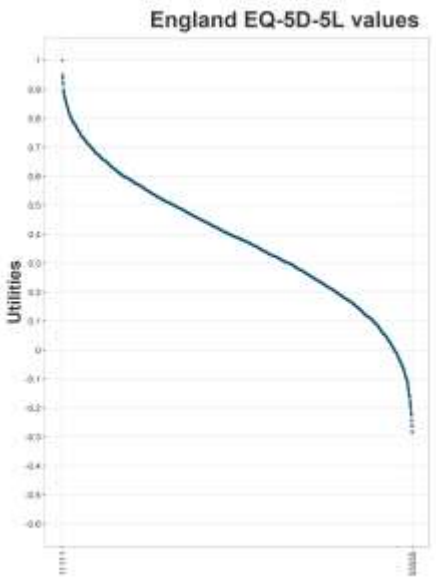
Issue: Recurring study (e.g. population observation) Concern: -3L used previously

- Recommendation:
  - Switch to the -5L.
  - With continued use of the -3L the problem will reoccur in the future. If you have to switch eventually, there is no better time than now.
  - Consider including both versions once or twice to establish a solid basis for linking previous and future data.

Changing value sets

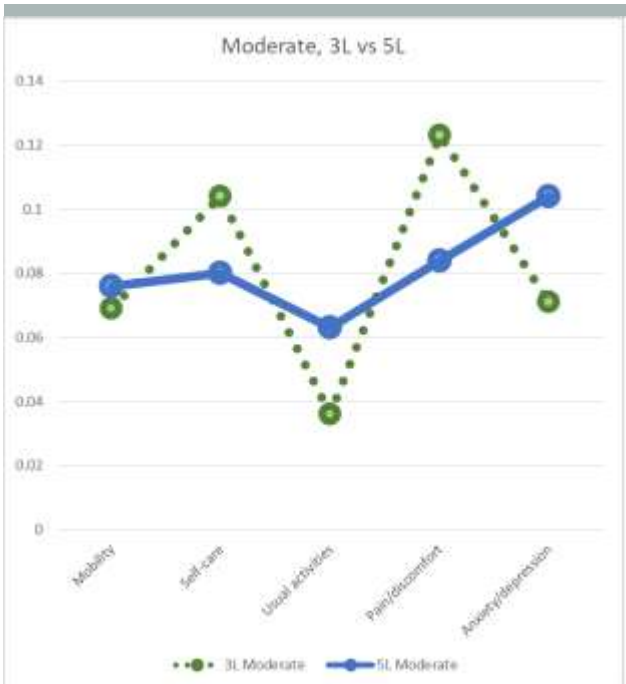
- Implies different estimates of QALYs gained, cost/QALY, and ICERs.
- General factors in play when changing value sets (regardless of version)
  - Altered slope
  - Altered relative ranking of states
  - Changes in “break points”
- Interactions between changes in descriptive system and value sets when moving from -3L to -5L:
  - More fine-grained descriptive system means fewer 11111 responses
  - Value difference of minimal change is reduced

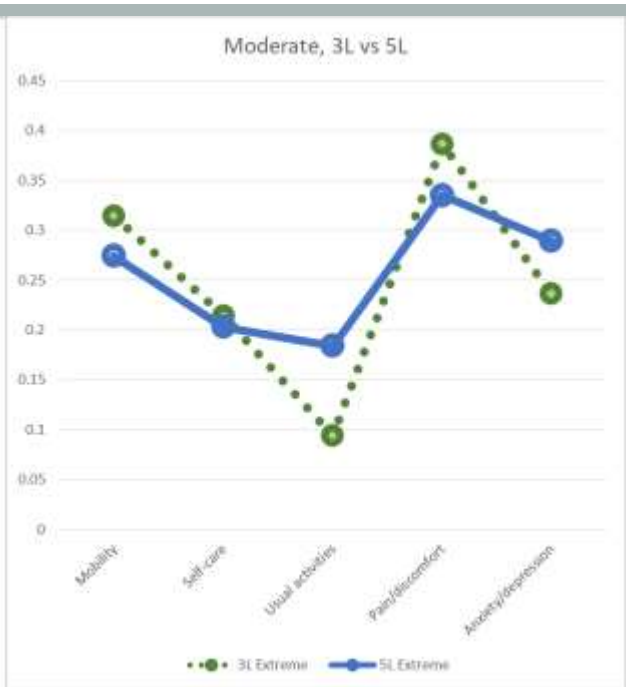


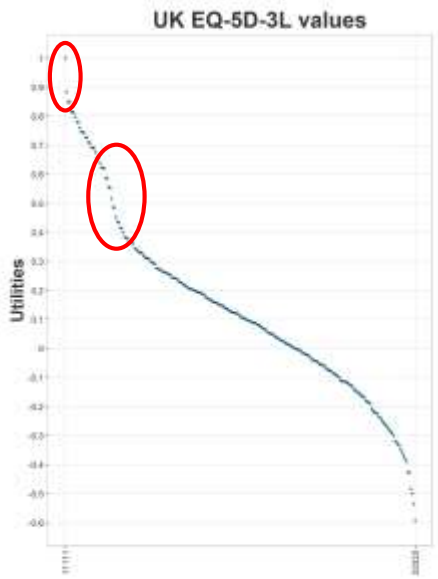


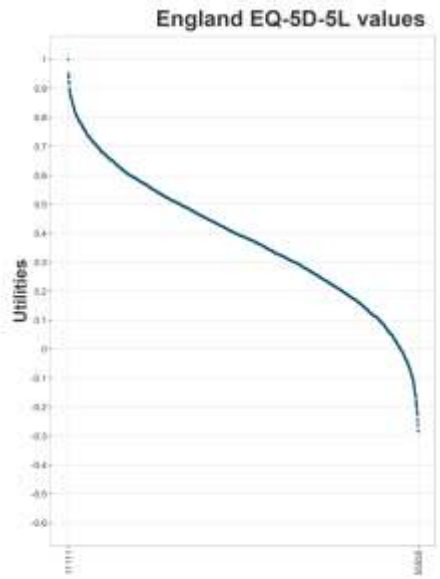
ge
ea
. ex
es t
ite
tha
e st
, th
e s

of life.
/ ranking of
itions or
ER
less steep
, since the
may be
ates for QOL-









Q-5D

int

t, th

ther

g at l

111 a

few

t bet

uities:

d those

aller in

nd, they

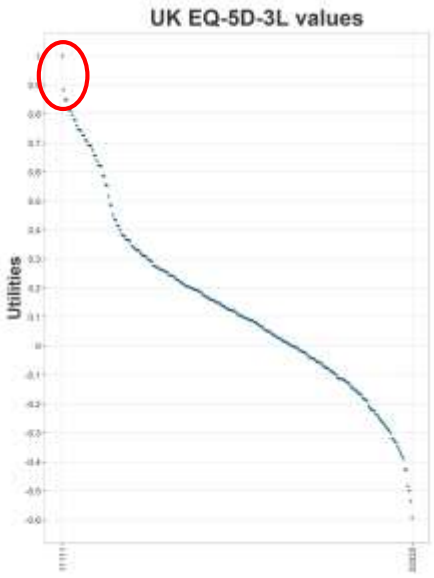
9-9-2018

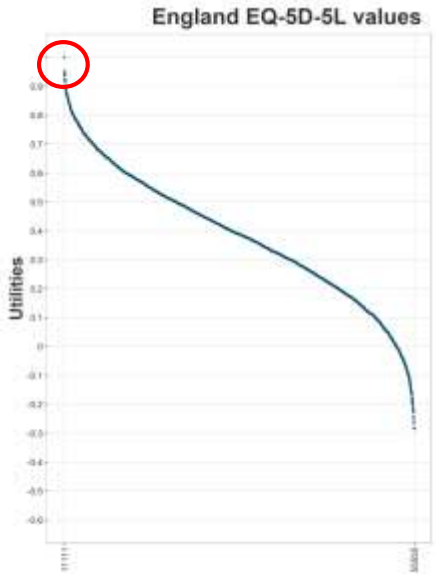
Changes in break points cont'd

- Break points produce areas with disproportionate potential QALY gains (and losses)
- For interventions that shift even a few individuals over the specific threshold, attenuating these break-points will result in smaller QALY gain estimates.
- While such breaks may seem somewhat implausible, and are likely caused by the highly granular nature of the descriptive system (particularly the -3L), this may be seen as a problem by certain users.

Interactions: Fewer 11111 responses with the -5L

- Since it is possible to describe smaller health problems with the -5L than the -3L, less respondents are placing themselves in the 11111 category
- As the value difference between 11111 and any other health state is substantial (even more so with the -3L), interventions that shifted a few respondents from states like 11211 to 11111 now produce fewer QALYs
  - Many respondents moving from -3L states like 11211 to 11111 will now move from -5L states like 11311 to 11211, for a much smaller QALY gain
  - Respondents moving from -5L state 11211 to 11111 will have a smaller QALY gain than -3L 11211 to 11111





9-9-2018

Interactions: minimum value differences reduced

- Respondents experiencing changes in health may be motivated to respond in such a way as to indicate change when administered questionnaires.
- With the new levels “slight” and “severe” in the -5L, the smallest possible changes are much smaller than in -3L
- QALY gains from reported small changes are thereby reduced with the introduction of the -5L.

Issues, summarized

- For reasons described previously, there are cases in which the -5L is likely to produce lower QALY estimates than the -3L
- When two value sets are in play, they will produce differences in QALY estimates
- If more than one value set is allowed by regulatory agencies, gaming is possible
- These issues are not caused by problems with the -5L, but by necessary consequences of greater resolution combined with new (improved) valuation studies

Suggested strategies for resolution

1. Regulatory agencies should not allow more than one value set even if they allow both versions
   - Where a -3L value set is preferred, -5L studies should use a -3L to -5L cross-walk
   - Where a -5L value set is preferred, -3L studies should use a -5L to -3L cross-walk
   - This should minimize the potential for gaming, and ensure an even playing field
2. Changes in value set should not be retroactive: old approvals should remain. This should reduce potential opposition to change from industry.
3. Future, challenging interventions should be considered using the new regime
4. New value sets should be set to a high standard of quality, and warrant scrutiny before acceptance.
5. When a high-quality value set for the -5L is available, regulatory agencies should change from -3L to -5L.