

GOING BEYOND HEALTH – THE DEVELOPMENT OF A NEW INTERNATIONAL GENERIC MEASURE THAT REFLECTS THE IMPACT OF TREATMENTS ON PATIENT AND CARER QUALITY OF LIFE

EXTENDING THE QALY PROJECT (E-QALY PROJECT)

This work has been funded jointly by the UK Medical Research Council and the EuroQoL Research Foundation. The authors have no financial conflict of interest.

ISPOR, Orlando (Virtually), Monday, May 18, 2020, 2:00 PM - 3:00 PM (EDT)

The E-QALY project

- Rationale and overview of EQALY project.
 - **John Brazier PhD**, Professor of Health Economics, ScHARR, University of Sheffield, UK.
- Challenges of developing instruments with international relevance.
 - **Federico Augustovski, PhD** Professor of Public Health, University of Buenos Aires, Argentina
- Psychometric evidence underpinning selection of items
 - **Clara Mukuria PhD**, Research Fellow, ScHARR, University of Sheffield, UK.
- Developing a new measure that meets the need of industry/policy makers
 - **A. Simon Pickard, PhD**, Professor of Pharmacy Systems, Outcomes and Policy, University of Illinois at Chicago, IL, USA.



Aim

- Identify a shorter list of items/questions for (1) a long measure and (2) a classification system based on psychometric performance

Which questions, populations and methods would address this aim?

Methods: questions

- 62 items from face validity results. Limited to best performing due to survey length. All domains/sub-domains apart from 'dignity'
- EQ-5D-5L, EQ-5D-3L, Short Warwick-Edinburgh Mental Wellbeing Scale (SWEMWBS) and the Adults Social Care Outcomes Toolkit (ASCOT) as well as sociodemographic questions.
- Three versions with different ordering of E-QALY items and EQ-5D-5L/3L order (where applicable). SWEMWBS and ASCOT always last
- Surveys managed by a single UK based company (Accent). Local ethics obtained for each country.

Methods: populations

- Patients, informal carers, social care users, general population
 - UK (n=2000) recruited online and from National Health Service (NHS) hospitals and primary care (paper, self complete)
 - Argentina (Spanish), Australia, China (Chinese), Germany (German), USA (n=500 to 900) recruited online
- All participants received incentives for taking part

Methods: analysis

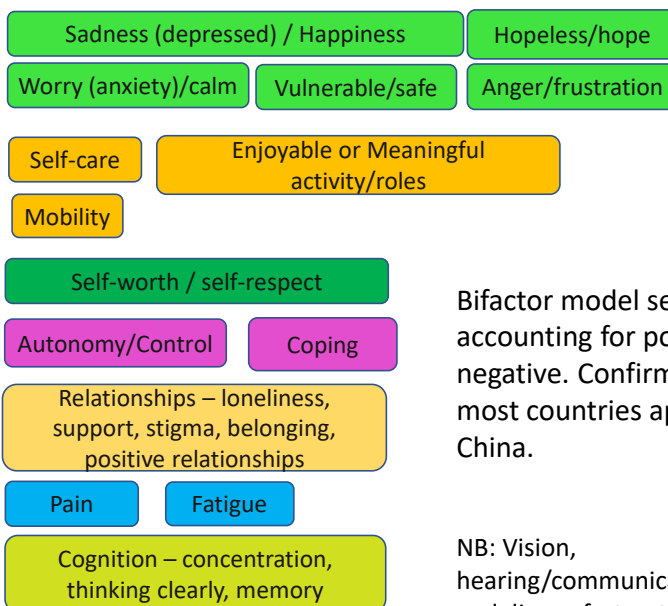
- Factor analysis to confirm domain structure – bifactor model to account for positive and negative items. UK first then applied to other countries
- Classical psychometric analysis (distributions of each item, missing items, known group differences)
- Item response theory (IRT) on separate domains/sub-domains – using graded response model (item fit, ordering of levels, differential item functioning (DIF)).
- Separate analysis with standardised protocol across the 6 countries

Results: samples

Country	N	Populations
Argentina	497	LT condition (64%); Carer (68%); Social Care (58%)
Australia	514	LT condition (73%); Health care aid user (57%); Carer (22%);
China	881	LT condition (72%); Carer (46%)
Germany	496	Healthy people (20%); Cancer (40%); Carer (40%)
UK	1,923	LT condition (76%); Carer (31%); Social Care (19%)
USA	903	LT condition (69%); Carer (22%); Social Care (10%)

LT – long term

Extending the QALY instrument sub-domains - profile



Bifactor model separately accounting for positive and negative. Confirmed in most countries apart from China.

NB: Vision, hearing/communication, sleep and discomfort not included

Results: classical psychometrics

- High proportion ($\geq 50\%$) report no problems in questions related to self-care, vision, hearing
- Low levels of missing data in the UK paper version
- Large to moderate effect sizes across the items for physical and mental health conditions; small effect sizes for number of hours spent caring.
- Personal needs/self-care questions had small effect sizes

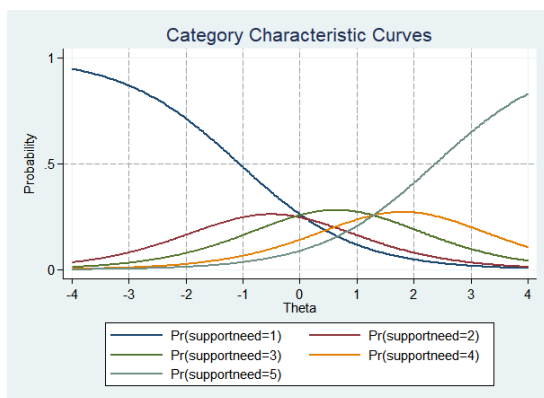
Generally difficult to identify poor performing items on the basis of classical psychometric results

Results: IRT

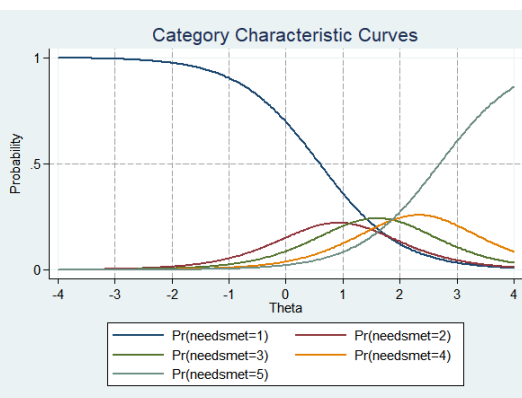
- Most items were in the expected sub-domain but some were better suited in other sub-domain e.g. *"I was able to do the things I wanted to do"* was in 'meaningful/enjoyable activities' but fitted in 'control'
- Item fit was poor in domains/sub-domains where the items were very similar e.g. pain, happiness, control
- There was evidence of DIF for some of the items
- Most of the items exhibited good range and good ordering of levels

Ordering problem examples – too many levels

I had support when I needed it



My personal needs were met



Discussion

- Some of the self-care/ personal needs questions consistently performed poorly
- Majority of the items performed well in terms of classical psychometrics
- IRT assessment mixed mainly due to DIF and item fit

Strengths

- Data drawn from multiple populations in multiple countries
- English, Chinese, German and Spanish tested
- Large overall sample to support decisions
- Application of best practice

Limitations

- Mostly online data with exception of sample from UK (n=627)
- No assessment of responsiveness to change over time and social care use
- IRT on single domains; no accounting for positive/negative

Summary

- Results from psychometric analysis used to provide evidence on the 62 items across 6 countries in English, Chinese, German and Spanish
- Summarised alongside face validity findings to support consultation process to select the experimental 25 item measure and 9 item classifier; latter may change following valuation
- Working with EQ group (IP holder) on distribution of measures
- Further data required to assess performance

Thanks

C.Mukuria@Sheffield.ac.uk