

# Does Assessing Outcomes in Terms of Capability for Schizophrenic Patients with Depression Provide more Information Than use of the NICE Recommended QALY? - An Empirical Comparison of the OxCAP-MH, ICECAP-A and EQ-5D-5L Instruments

Timea Mariann Helter (✉ [timea.helter@muv.ac.at](mailto:timea.helter@muv.ac.at))

Medical University of Vienna <https://orcid.org/0000-0002-2696-2553>

Joanna Coast

University of Bristol Medical School

Agata Łaszewska

Medizinische Universität Wien

Tanja Stamm

Medizinische Universität Wien

Judit Simon

Medizinische Universität Wien

---

## Research

**Keywords:** capability approach, psychometric validation, OxCAP-MH, ICECAP-A, EQ-5D-5L, schizophrenia, depression, mental health, quality of life, wellbeing, UK

**Posted Date:** July 23rd, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-44326/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

## Background

There is increasing evidence that assessing outcomes in terms of capability wellbeing provides information beyond that of health-related quality of life measures for evaluation in mental health research. This paper aims to comprehensively compare the properties of the Oxford CAPabilities questionnaire-Mental Health (OxCAP-MH), the ICECAP-A the EQ-5D-5L descriptive system and EQ-5D VAS in schizophrenic patients with depression.

## Methods

Using trial data for 100 patients from the UK, the properties of the instruments were compared in terms of construct validity, including correlations between the OxCAP-MH, the ICECAP-A, the EQ-5D-5L descriptive system and the EQ-5D VAS scores; and comparative assessment of their sensitivity to change based on external anchors. Exploratory factor analysis (EFA) investigated the extent to which the instruments measure complementary or overlapping constructs. The pattern and extent of agreement between all instruments was plotted on Bland-Altman diagrams.

## Results

Different aspects of the analysis confirmed that the capability instruments had stronger convergent validity with each other than with health-related instruments. The EFA found that while the EQ-5D-5L descriptive system loads onto one factor, the items of the ICECAP-A load onto three factors and the items of the OxCAP-MH spread across four factors. Correlation between the OxCAP-MH and ICECAP-A change scores was moderate (0.389). The ICECAP-A change scores also moderately correlated with change scores of generic health-related scales (0.307-0.357) and disease-specific instruments (0.295-0.468). The OxCAP-MH change scores had low correlation to generic (0.153-0.202) and moderate to high correlation with disease-specific instruments (0.441-0.527). The Bland Altman plots showed small average discrepancy between the four scales. However, the limits of agreement were wider and therefore more ambiguous in the comparison between the EQ-5D-5L descriptive system score and the capabilities instruments than in the direct comparison of OxCAP-MH and ICECAP-A.

## Conclusions

Assessing outcomes in terms of capability for schizophrenic patients with depression provide more information than use of the NICE recommended QALY. OxCAP-MH and ICECAP-A show similar construct validity in severely ill mental health patients within the capability framework. Future research should extend the comparison of the properties of these instruments to other areas of mental health.

## Introduction

The capability approach was developed by Amartya Sen with a core focus on what individuals are free and able to do (i.e., capable of) (1, 2). This approach places emphasis on promoting wellbeing through enabling people to realise their capabilities and engage in behaviours that they value (3). There is increasing interest in the use of the capability approach for the economic evaluation of health-related interventions (4). One reason for this is the wider evaluative space this approach offers in comparison to the commonly used methods of assessment (5). Quality-Adjusted Life Years (QALYs) are routinely used as a summary measure of health outcome for economic evaluation, which incorporates the impact on both the quantity and quality of life (6). The quality component is measured with preference-based utility values of health-related quality of life (HRQoL) instruments. Currently EQ-5D is the most commonly recommended such instrument in a number of settings, including the National Institute for Health and Care Excellence (NICE) in the UK (7, 8). In its current form, however, QALYs may not capture important consequences where impacts of interventions go beyond a rather narrow definition of health. For instance, QALYs may be insensitive to the impact of social care interventions and therefore underestimate their full welfare impact in the area of mental health (9). Mental health care interventions usually target both health and social impairments because many people with severe and enduring mental illness experience significant functional and social challenges (10).

A recent literature review of capability instruments in economic evaluations of health-related interventions has identified 14 instruments, differing in their domains, levels, target populations and interventions (4). Two of these instruments are commonly used and have been validated for the adult population with mental health problems: the Oxford CAPabilities questionnaire-Mental Health (OxCAP-MH) and the ICECAP measure for Adults (ICECAP-A). Both instruments have been shown to move beyond the standard HRQoL approach for the measurement and valuation of outcomes (5, 10–18). While both instruments are grounded in the capability approach and have been implemented in the mental health context, their conceptual approaches differ. The OxCAP-MH is rooted in Nussbaum's central human capabilities and was developed free from geographical and cultural contexts. It was published in 2013 (10). The ICECAP-A belongs to a broader group of ICECAP capability instruments, each focusing on different aspects of capabilities and life span. It draws on the capability approach, using participatory (qualitative) methods to generate attributes as recommended by Sen (19). The ICECAP-A descriptive system was published in 2012.

Questions remain about whether different applications of the same broad concept of the capability approach result in similar or different measurement properties. Comparative studies of the measurement properties of alternative capability instruments have not been conducted yet, and researchers cannot rely on published studies when choosing between instruments. The lack of such comparative information hinders the future optimisation of research efforts related to quality of life and wellbeing in the (mental) health field.

Exploring the construct structure and the convergence and divergence between the ICECAP-A and the OxCAP-MH measures would not only contribute to our understanding of which measure may be used in certain settings and provide further information about their complementary or enhanced conceptual properties, but it may also shed light on some broader questions about how each method of instrumenting the capability theory influences measurement processes. Moreover, the hypothesis that capability instruments, even when derived from differing conceptual underpinnings, are more correlated to each other than to a HRQoL instrument, e.g. EQ-5D-5L or EQ-5D VAS, has not been tested before in the area of mental health. This paper aims to contribute to the

utilisation of the capability approach in mental health research, by exploring the empirical relationship between the OxCAP-MH, the ICECAP-A and the EQ-5D instruments. More specifically, the purpose of this study is to examine correlations between the OxCAP-MH, the ICECAP-A, the EQ-5D-5L descriptive system and the EQ-5D VAS scores, explore whether they measure complementary or overlapping constructs, and investigate the similarities in how they capture change. The focus of the paper is on the comparability of the descriptive systems of the instruments, therefore, preference-based weights that are available for the EQ-5D-5L and the ICECAP-A at the time of writing this paper were not used. Moreover, relevant tariff values for the EQ-5D-5L descriptive system and the ICECAP-A have different anchor points. The 0 point of the EQ-5D-5L value set is anchored against 'death', while the 0 point of the ICECAP-A value set is anchored against 'no capability' leading to potential difficulties in interpreting in any comparisons based on preference-weighted scales.

## Methods

### Data source

The analysis in this paper was based on data from the PoMeT trial (20), which investigated the impact of Positive Memory Training on depression symptoms of schizophrenia patients ( $n = 100$ ) in the UK between 2014–2016. The trial received ethical approval from the Berkshire Research Ethics Committee (REC ref 13/SC/0634). Patients were eligible for inclusion if they were between 18–65 years of age, had a DSM-V diagnosis of schizophrenia or schizoaffective disorder, and had at least a mild level of depression as measured by scoring 14 or more on the Beck Depression Inventory-II (21). Patients were assessed at four time points through the 9-month study period: baseline, 3 months, 6 months and 9 months. More details about the PoMeT trial can be found in Steel et al (20).

### Instruments

The OxCAP-MH is a self-reported, 16-item, mental health specific instrument, where items are rated on a 1–5 Likert-scale and each question provides an equal contribution to the overall score. The 16 items cover a broad range of individual wellbeing including: Overall health, Enjoying social and recreational activities, Losing sleep over worry, Friendship and support, Having suitable accommodation, Feeling safe, Likelihood of discrimination and assault, Freedom of personal and artistic expression, Appreciation of nature, Self-determination and Access to interesting activities or employment (10). The OxCAP-MH initial score (16–80 scale) is converted on to a 0–100 scale referring to minimum and maximum capabilities using the formula:  $100 \times (\text{OxCAP-MH total score} - \text{minimum possible score}) / \text{possible range}$  (11). Higher scores indicate better capabilities; items 2, 4, 5, 6, 9, 10, 11, 12, 13, 14, 15 and 16 are reverse coded. The OxCAP-MH has shown validity (5, 11), responsiveness (5, 11) and feasibility (10) in several settings and mental health disease areas and is currently available in the English, German (22) and Hungarian (23) languages with further language translations ongoing. In an earlier factor analysis, Laszewska et al. found that all EQ-5D-5L items and seven OxCAP-MH items loaded on one factor and nine remaining OxCAP-MH items loaded on a separate factor, indicating that the OxCAP-MH may be seen as supplementary rather than complementary in its concept, when compared to the EQ-5D-5L (5). The OxCAP-MH does not yet have a preference-based value set; however, research is on-going to develop a weighting system for its domains.

The ICECAP-A is a brief self-reported measure for the general adult population with five items, each of which can take one of four levels ranging from full capability to no capability. The domains include Stability (being able to feel settled and secure), Attachment (being able to have love, friendship and support), Autonomy (being able to be independent), Achievement (being able to achieve and progress), and Enjoyment (being able to have enjoyment and pleasure) (12). The ICECAP-A has shown validity (16, 17, 19, 24, 25) reliability (26, 27), responsiveness (28) and feasibility (14) in different populations. Beside the original English language version, it is also available in German (26), Chinese (29), Welsh, Dutch, Danish, Persian and Italian languages (30). Previous factor analysis comparing the ICECAP-A with the items of EQ-5D-5L (31) and EQ-5D-3L (13, 15) found that these instruments measure two different constructs and therefore provide potentially different information. A recent systematic literature review found inconsistencies between the ICECAP-A and EQ-5D instruments, suggesting that the ICECAP-A is most appropriately regarded as a complement for and not a substitute to the EQ-5D-3L and EQ-5D-5L in particular (32). The ICECAP-A has a preference-based value set derived from the UK general population (24) and it is increasingly used in economic evaluations (32). The simple addition of ICECAP-A level sum scores ranges from 5 to 20, with higher scores representing better capabilities.

The EQ-5D-5L is one of the most commonly used self-reported generic health status measures, and its validity and reliability have been reported in various health conditions and populations (33). The EQ-5D-5L descriptive system comprises five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Beside the original 3-level version (34), a more sensitive, 5-level version exists since 2009 (35). Both versions have value sets in several countries (36); but they can also be used as simple descriptive systems with total scores ranging from 5–15 for the 3L version and 5–25 for the 5L version, with higher scores representing better HRQoL. As part of this instrument, respondents' self-rated health is also recorded on a vertical visual analogue scale (EQ-5D VAS) where scores range between 0–100 referring to worst imaginable health state and best imaginable health state, respectively.

Since the OxCAP-MH and EQ-5D VAS scores range between 0–100, the ICECAP-A level sum scores range between 5–20 and the EQ-5D-5L descriptive system level sum scores range between 5–25, the comparisons between the instruments would be challenging. Hence, all values were transformed to a 0–1 range for the relevant statistical calculations, i.e. in case of responsiveness and agreement analysis. This was calculated as a simple division by 100 in case of the OxCAP-MH and EQ-5D VAS scores, and a transformation of the ICECAP-A and EQ-5D-5L scores in a way that a score of 5 was recalibrated to 0 and scores of 20 and 25 were recalibrated to 1, respectively.

The Beck Depression Inventory (BDI), General Anxiety Disorder (GAD), Rosenberg self-esteem scale (RSES), and the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) are all mental-health specific, self-reported outcome instruments. They were used as anchors for the sensitivity to change analysis to assess external responsiveness.

BDI is a self-reported measure of depressive symptoms and their severity in adolescents and adults according to the Diagnostic and Statistical Manual for Mental Disorder (37). It has 21-items scored on 4-point polytomous response scale ranging from 0 to 3 (21). Scores range between 0 and 63 with higher score

representing more severe depression.

GAD is a self-reported measure of anxiety symptoms over the last two weeks. It consists of seven items scored on a 0–3 scale with higher score indicating more severe symptoms (range from 0 to 21) (38). The cut-off scores of 5, 10 and 15 reflect mild, moderate and severe anxiety symptoms, respectively (39).

RSES is a 10-item, self-reported instrument that measures global self-worth by measuring both positive and negative feelings about the self (40). Items are answered using a 4-point polytomous response scale format ranging from strongly agree to strongly disagree. Items 2, 5, 6, 8, 9 are reverse scored.

The self-reported WEMWBS instrument was developed in the UK to assess mental wellbeing including affective-emotional aspects, cognitive-evaluative dimensions and psychological functioning. It is a 14-item scale with 5 response categories ('none of the time', 'rarely', 'some of the time', 'often', 'all of the time'), with a total score ranging from 14–70. A higher score indicates a higher level of mental wellbeing (41).

## Statistical analysis

The statistical analysis focused on exploring and comparing the measurement properties of the OxCAP-MH, ICECAP-A, EQ-5D-5L descriptive system and the EQ-5D VAS. Exploratory factor analysis (EFA), correlations of baseline and change scores to test and compare construct validity across the scales, and investigation of responsiveness to change and degree of agreement were carried out.

For all analyses, the level of significance was determined at  $p < 0.05$ , unless stated otherwise. Group comparisons of mean baseline scores were conducted using t-tests for two-group comparisons and ANOVA for multiple group comparison. Analysis was conducted on complete cases, excluding missing items at the relevant time point, unless stated otherwise. EFA was conducted with the freely available FACTOR software, and we used STATA Version 16 for all other analyses.

## Construct validity

Construct validity indicates the degree to which the scores of the capability and HRQoL instruments are consistent with the underlying concepts of these wellbeing measures (42, 43). Graphical presentation of correlation between baseline and change scores explored the degree of agreement between the four scales. The axis of the graphs represented the minimum and maximum values of the relevant instruments. The hypothesis, that capability instruments and their items have stronger correlation with each other than with a HRQoL instrument, was tested through exploring the correlation between baseline scores. Pearson correlations across OxCAP-MH, ICECAP-A, EQ-5D-5L descriptive system and EQ-5D VAS were calculated at total score-level and at item-level for each time point and assessed based on Cohen's effect size classification, namely  $< 0.3$  is small,  $0.3 - < 0.5$  is moderate and  $\geq 0.50$  is large (44).

## Exploratory factor analysis

EFA was conducted on the baseline scores of the OxCAP-MH, ICECAP-A and EQ-5D-5L to examine the overlap between the constructs of the two capability measures and the multidimensional measure of HRQoL, and to study how far they share the same set of underlying factors. Further details on the methods of EFA can be found in Appendix 1.

## Responsiveness

Responsiveness was defined as the ability to capture clinically important changes over time (45). Patients filled out each four scales at both baseline and 9 months, which allowed for an exploration of change in mean scores over time. Responsiveness was assessed in terms of an external approach comparing the extent to which change in a capability measure relates to corresponding change in anchor instruments (46, 47). The analysis of responsiveness started with the definition of 2–4 instruments which could be used as autonomous anchors because they identify change that is unlikely to have arisen by chance (47).

The level of responsiveness was evaluated by defining groups who worsened, improved or remained stable, based on whether a change in the instrument scores between baseline and 9-month follow-up assessments was measured for individuals by the reference or anchor instruments. The calculation was based on the difference between baseline to 9-month values of standard error of measurement (SEM) using the following formula:

$$S_{diff} = \sqrt{(SEM_1^2 + SEM_2^2)}$$

SEM was calculated by using the standard deviation (SD) of the instrument multiplied by the square root of one minus its reliability coefficient at baseline and 9 months (11, 48). Internal consistency reliability coefficients were calculated for each scale based on the baseline to 3-month and 6-month to 9-month follow-up scores. More details on the calculation of the difference in SEM values can be found in Appendix 5. There is no consensus about how many SEMs an individual's score must change for that change to be considered clinically meaningful. This paper used the threshold of one SEM, which is known to frequently correspond to a minimally important difference (11, 49). In addition, standardised response mean (SRM) was calculated as the ratio of the mean change, between baseline and 9-month follow-up scores in a single group, to the SD of the change scores (42). Small, moderate and large magnitude of change was indicated by 0.20–0.49, 0.50–0.79 and  $\geq 0.80$  values of SRM, respectively (33). Next, the percentages of the study respondents who improved, worsened or remained stable according to the capability and anchor questionnaires were calculated to explore changes at the individual patient level (5).

## Agreement analysis

The pattern and extent of the agreement between OxCAP-MH, ICECAP-A, EQ-5D-5L descriptive system and EQ-5D VAS scores were plotted on Bland and Altman diagrams (50), where the difference between the instruments is shown on the vertical axis of the diagram against the mean of the pair on the horizontal axis (51).

## Results

# Patient characteristics

Patient characteristics are presented in Table 1.

Table 1  
Patient characteristics and mean baseline OxCAP-MH, ICECAP-A, EQ-5D-5L descriptive system and EQ-5D VAS score:

Scale (Min-Max)		Overall data		OxCAP-MH (0-100)			ICECAP-A (5-20)				EQ-5D-5L descriptive system (25)			f	
		N	% or mean (SD)	n	Mean baseline score (SD)	0-1 score*	p**	n	Mean baseline score (SD)	0-1 score*	p**	n	Mean baseline score (SD)		0-1 score*
Age	[Overall sample]	100	42.97 (10.55)	93	55.66 (12.90)	0.56		97	11.35 (2.86)	0.42		100	18.73 (4.03)	0.69	
Gender	Male	75	75%	69	56.46 (12.25)	0.56	0.318	72	11.43 (3.03)	0.43	0.642	75	18.89 (0.45)	0.69	
	Female	25	25%	24	53.39 (14.65)	0.53		25	11.12 (2.32)	0.41		25	18.24 (0.88)	0.66	
Higher education	Yes	49	49%	46	53.36 (1.93)	0.53	0.089	47	11.04 (2.63)	0.40	0.306	49	18.45 (0.63)	0.67	
	No	51	51%	47	57.91 (1.81)	0.58		50	11.64 (3.06)	0.44		51	19.00 (0.51)	0.70	
Living situation	Living with family	19	19%	19	51.48 (13.20)	0.51	0.340	19	10.84 (2.27)	0.39	0.799	19	19.58 (2.91)	0.73	
	Renting a flat	12	12%	10	59.06 (16.38)	0.59		10	11.20 (2.86)	0.41		12	19.00 (3.59)	0.70	
	Owning a flat	5	5%	4	51.95 (5.90)	0.52		5	11.00 (1.73)	0.40		5	16.60 (4.22)	0.58	
	Other	64	64%	60	56.67 (12.41)	0.57		63	11.56 (3.10)	0.44		64	18.59 (4.37)	0.68	
Employment	Employed full-time	3	3%	3	67.19 (14.32)	0.67	0.153	3	13.33 (1.53)	0.56	0.609	3	19.00 (3.46)	0.70	
	Employed part-time	9	9%	8	59.18 (9.21)	0.59		8	11.75 (1.39)	0.45		9	19.44 (4.00)	0.72	
	Unemployed	86	86%	80	55.22 (13.00)	0.55		84	11.26 (3.01)	0.42		86	18.86 (3.88)	0.69	
	Other (Student/Retired)	2	2%	2	42.19 (8.84)	0.42		2	10.50 (0.71)	0.37		2	9.50 (0.71)	0.23	
Primary diagnosis	Schizophrenia	70	70%	66	56.72 (12.77)	0.57	0.216	68	11.34 (2.89)	0.42	0.949	70	18.84 (0.50)	0.69	
	Schizoaffective or psychosis NOS	30	30%	27	53.07 (2.52)	0.53		29	11.38 (2.81)	0.43		30	18.47 (0.69)	0.67	
Depression severity	Mild/moderate	41	41%	40	63.44 (1.58)	0.63	0.000	41	12.88 (2.25)	0.53	0.000	41	20.07 (0.58)	0.75	
	High	59	59%	53	49.80 (1.61)	0.50		56	10.23 (2.75)	0.35		59	17.80 (0.52)	0.64	
Intervention	Treatment	49	49%	46	55.10 (12.59)	0.55	0.678	49	11.33 (2.76)	0.42	0.934	51	18.24 (0.53)	0.66	
	Control group	51	51%	47	56.22 (13.30)	0.56		48	11.36 (2.99)	0.42		49	19.25 (0.61)	0.71	

\*scores were standardised to a 0-1 range for all instruments for reasons of comparability

\*\*t test for two-group comparison, ANOVA for multiple group comparison

(Insert Table 1)

Mean baseline scores for all instruments used in this analysis are presented in Table 2.

Table 2  
Baseline scores of the relevant patient-reported outcome measures used in the trial

	N	Mean	SD	Min	Max	Theoretical Min-Max
OxCAP-MH	93	55.66	12.90	21.9	87.5	0-100
ICECAP-A	97	11.35	2.86	6	19	5–20
EQ-5D-5L descriptive system	100	18.73	4.03	6	25	5–25
EQ-5D VAS	99	50.14	21.19	3	95	0-100
BDI	100	30.45	9.99	14	52	0–63
GAD	100	11.16	5.62	0	21	0–21
RSES	100	28.50	5.48	14	40	10–40
WEMWBS	100	34.25	10.54	14	65	14–70

(insert Table 2)

## Construct validity

Both the graphical (Fig. 1) and numerical (Table 3) presentation of correlations at baseline confirmed the hypothesis that the capability instruments are more correlated with each other than with the EQ-5D-5L descriptive system's level sum scores or the EQ-5D VAS. Correlations between the capability and HRQoL measures (0.315–0.385) were lower than those between OxCAP-MH and ICECAP-A (0.641). The ICECAP-A was slightly more correlated with EQ-5D VAS (0.385) than with the EQ-5D-5L descriptive system (0.354), whilst the OxCAP-MH was somewhat higher correlated with the EQ-5D-5L descriptive system (0.370) than with the EQ-5D VAS (0.315).

Table 3  
Pearson correlations between OxCAP-MH, ICECAP-A, EQ-5D-5L index and EQ-5D VAS baseline scores

	ICECAP-A	EQ-5D-5L descriptive system	EQ-5D VAS
OxCAP-MH	<b>0.641**</b> (n = 92)	<i>0.370**</i> (n = 93)	<i>0.315**</i> (n = 93)
ICECAP-A		<i>0.354**</i> (n = 97)	<i>0.385**</i> (n = 97)
EQ-5D-5L*			<b>0.509</b> (n = 99)
** $p < 0.01$ , Moderate correlations (0.3–0.5) in italic, Strong correlations ( $\geq 0.5$ ) in bold			

(insert Fig. 1)

(insert Table 3)

## Exploratory factor analysis

A four-factor solution was chosen according to the Kaisers criterion based on a scree plot, as described in the Appendix. EFA with four factors found that all items of the instruments had communalities greater than 0.35, i.e. none of the items struggled to load significantly on any factor. Hence, the factor loadings are shown in Table 4 for any factor  $> 0.35$ .

Table 4  
Exploratory factor analysis of the OxCAP-MH, ICECAP-A and EQ-5D-5L items with 4 factors using  
promin rotation

	Factor 1	Factor 2	Factor 3	Factor 4
<b>OxCAP-MH</b>				
Daily activities	0.453			
Social networks			0.646	
Losing sleep		0.523		
Enjoy recreation			0.584	
Suitable accommodation	0.391			
Neighbourhood safety		0.411		
Potential for assault		0.513		
Discrimination		0.635		
Influence local decisions			0.517	
Freedom of expression			0.384	
Appreciate nature				0.688
Respect and appreciation of people				0.786
Love and support		0.381	0.562	
Planning one's life		0.510	0.494	
Imagination and creativity			0.498	
Access to interesting activities		-0.513	0.752	
<b>ICECAP-A</b>				
Feeling settled and secure		0.499	0.404	
Love, friendship and support			0.725	
Being independent	0.385			
Achievement and progress			0.598	
Enjoyment and pleasure			0.812	
<b>EQ-5D-5L</b>				
Mobility	0.928			
Self-care	0.806			
Usual activities	0.754			
Pain	0.761			
Anxiety and depression	0.389			
<i>Loadings <math>\leq 0.35</math> were removed</i>				

(insert Table 4)

Factor one consisted of the five EQ-5D-5L descriptive system domains, with particularly high communalities for all items apart from the *Anxiety and depression* domain. The *Daily activities* and *Suitable accommodation* domains of OxCAP-MH and the *Being independent* domain of ICECAP-A also loaded to this undoubtedly physical health related factor.

None of the EQ-5D-5L descriptive system domains loaded on factors two, three and four. Only the *Feeling settled and secure* domain of ICECAP-A loaded on factor two, where high communalities were observed for the domains of OxCAP-MH related to the perception of the settlement and security, e.g. *Losing sleep*, *Neighbourhood safety*, *Potential for assault* and *Discrimination*. The negative loading of *Access to interesting activities* is consistent with the direction of scoring of the items.

Factor three consisted of four ICECAP-A domains (the *Being independent* domain did not load on this factor) and the *Social networks*, *Enjoy recreation*, *Influence local decisions*, *Freedom of expression*, *Love and Support* and *Planning one's life* domains, *Imagination and creativity* and *Access to interesting activities* domains of OxCAP-MH.

Factor four consisted of two OxCAP-MH domains, both focusing on the appreciation of a person's environment. These two domains had remarkably high communalities on factor four and did not load to any other factor.

## Responsiveness

The Pearson correlation between the baseline to endpoint change scores of the OxCAP-MH, ICECAP-A, EQ-5D-5L and EQ-5D VAS, and the potential reference instruments are presented in Table 5.

Table 5  
Pearson correlations between change scores of OxCAP-MH, ICECAP-A and EQ-5D-5L descriptive system and EQ-5D VAS scores

	OxCAP-MH	ICECAP-A	EQ-5D-5L	EQ-5D VAS
OxCAP-MH	1.000			
ICECAP-A	n = 78 <i>0.389</i>	1.000		
EQ-5D-5L	n = 79 0.202	n = 88 <i>0.357</i>	1.000	
EQ-5D VAS	n = 79 0.153	n = 88 <i>0.307</i>	n = 90 <i>0.429</i>	n = 90 1.000
Beck Depression Inventory (BDI)	n = 79 <i>-0.448</i>	n = 88 -0.295	n = 95 -0.193	n = 90 -0.179
Generalized Anxiety Disorder scale (GAD)	n = 79 <b>-0.527</b>	n = 88 <i>-0.417</i>	n = 90 <i>-0.325</i>	n = 90 -0.267
Rosenberg Self-Esteem Scale (RSES)	n = 79 <i>-0.441</i>	n = 87 <i>-0.355</i>	n = 90 -0.227	n = 90 -0.268
Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS)	n = 79 <b>0.521</b>	n = 88 <i>0.468</i>	n = 90 0.182	n = 90 0.271
<i>Moderate correlations (0.3–0.5) in italic, Strong correlations (&gt;= 0.5) in bold</i>				

(insert Table 5)

Correlation between the OxCAP-MH and ICECAP-A change scores was moderate (0.389). The ICECAP-A change scores also moderately correlated with change scores of generic health-related scales (0.307–0.357) and disease-specific instruments (0.295–0.468). The OxCAP-MH change scores had low correlation to generic (0.153–0.202) and moderate to high correlation with disease-specific instruments (0.441–0.527). Since the GAD and WEMWBS measures had the highest correlation with the four wellbeing instruments under investigation in this paper, they were selected as suitable reference anchor instruments for the analysis of responsiveness.

(insert Table 6)

Table 6  
Descriptive statistics (mean, SD) by external criteria (change defined by GAD and WEMWBS) using complete cases

Scale	Time point	Generalized Anxiety Disorder scale (GAD)						Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS)								
		Improved		Stable		Deteriorated		Improved		Stable						
		N	Original scores (SD)	0-1 scores (SD)	N	Original scores (SD)	0-1 scores (SD)	N	Original scores (SD)	0-1 scores (SD)	N	Original scores (SD)	0-1 scores (SD)	N	Original scores (SD)	0-1 scores (SD)
OxCAP-MH	Baseline	26	54.87 (12.65)	0.55 (0.13)	44	54.33 (13.52)	0.54 (0.14)	16	58.30 (9.80)	0.58 (0.10)	25	53.75 (14.23)	0.54 (0.14)	49	54.72 (12.18)	0.55 (0.12)
	9 months	26	66.11 (14.66)	0.66 (0.15)	44	54.97 (15.52)	0.55 (0.16)	15	47.40 (14.62)	0.47 (0.15)	27	61.52 (18.07)	0.62 (0.18)	48	55.60 (14.68)	0.56 (0.15)
	Change	24	10.48 (9.49)	0.10 (0.09)	41	0.84 (9.37)	0.01 (0.09)	14	-9.04 (12.40)	-0.09 (0.12)	24	8.27 (11.38)	0.08 (0.11)	45	1.11 (10.40)	0.01 (0.10)
	<i>p</i> value*		<b>0.000</b>		0.570		0.017		0.002		0.478					
SRM**	24	0.880		41	0.071		14	0.759		24	0.694		45	0.093		
ICECAP-A	Baseline	26	10.69 (1.98)	0.38 (0.13)	47	11.09 (3.01)	0.41 (0.20)	17	12.65 (3.22)	0.51 (0.21)	26	10.69 (2.72)	0.38 (0.18)	52	11.42 (2.96)	0.43 (0.20)
	9 months	27	13.00 (2.87)	0.53 (0.19)	47	11.28 (3.21)	0.42 (0.21)	15	12.00 (3.57)	0.47 (0.24)	27	13.26 (3.77)	0.55 (0.25)	51	11.31 (2.78)	0.42 (0.19)
	Change	26	2.27 (2.55)	0.15 (0.17)	47	0.19 (2.05)	0.01 (0.14)	15	-0.27 (3.01)	-0.02 (0.20)	26	2.54 (2.96)	0.17 (0.20)	51	-0.04 (2.05)	-0.00 (0.14)
	<i>p</i> value*		<b>0.009</b>		0.525		0.737		<b>0.000</b>		0.892					
SRM**	26	0.883		47	0.074		17	0.105		26	0.988		51	0.016		
EQ-5D-5L descriptive system	Baseline	28	18.93 (3.72)	0.70 (0.19)	47	18.30 (4.16)	0.66 (0.21)	17	18.24 (4.53)	0.66 (0.23)	28	18.89 (4.93)	0.69 (0.25)	52	18.13 (3.77)	0.66 (0.19)
	9 months	28	21.18 (2.60)	0.81 (0.13)	47	18.04 (4.42)	0.65 (0.22)	17	18.53 (3.95)	0.67 (0.20)	28	20.07 (4.52)	0.75 (0.23)	52	18.52 (3.95)	0.68 (0.20)
	Change	28	2.25 (2.80)	0.11 (0.14)	47	-0.26 (3.30)	-0.01 (0.16)	17	0.29 (3.20)	0.01 (0.16)	28	1.18 (3.21)	0.06 (0.16)	52	0.38 (3.41)	0.02 (0.17)
	<i>p</i> value*		<b>0.000</b>		0.600		0.709		0.063		0.420					
SRM**	28	0.369		47	0.043		17	0.048		28	0.193		52	0.062		
EQ-5D VAS	Baseline	27	50.93 (20.71)	0.51 (0.21)	47	48.59 (20.70)	0.49 (0.21)	17	49.41 (21.93)	0.49 (0.22)	27	46.94 (20.95)	0.47 (0.21)	52	50.04 (21.08)	0.50 (0.21)
	9 months	28	65.43 (18.88)	0.65 (0.19)	47	48.81 (23.63)	0.49 (0.24)	16	47.50 (22.36)	0.48 (0.22)	28	63.29 (22.03)	0.63 (0.22)	51	48.02 (21.96)	0.48 (0.22)
	Change	27	13.78 (19.90)	0.14 (0.20)	47	0.22 (21.28)	0.002 (0.21)	16	0.94 (21.85)	0.01 (0.22)	27	15.54 (16.36)	0.16 (0.16)	51	-1.14 (21.62)	-0.01 (0.22)
	<i>p</i> value*		<b>0.001</b>		0.943		0.866		<b>0.000</b>		0.709					
SRM**	27	3.118		47	0.050		16	0.213		27	3.516		51	0.258		

\**t*-test between baseline and 9 months scores of the relevant group;

\*\*SRM was calculated as the ratio of the mean change, between baseline and follow-up scores in a single group, to the SD of the change scores (OxCAP-MH: 2.57/0.17; EQ-5D-5L d.s.:6.10/0.17; VAS: 4.42/0.22)

Table 6 presents the number of patients improved, deteriorated and remained stable based on assessment by different anchors, and the mean scores in each group. Each instrument captured changes in patients' health state with somewhat similar magnitude. For the study participants who reported improvement in GAD, the improvements in the OxCAP-MH, ICECAP-A and EQ-5D VAS scores were statistically significant at the 1% level with large SRM statistics. However, improvement in WEMWBS was associated with statistically significant improvement at the 1% level with large SRM statistics reported only in case of the ICECAP-A and the EQ-5D VAS measures, with moderate results for OxCAP-MH. The effect sizes were lower for the EQ-5D-5L descriptive system.

(insert Table 7)

Table 7  
Number of patients improved, deteriorated or remained stable as defined by the investigated and anchor questionnaires (based on SEM)

Scale (complete cases)	Change*	N	GAD (n = 92)			WEMWBS (n = 91)			
			Improved	Stable	Deteriorated	N	Improved	Stable	Deteriorated
	N		28	47	17	28	52	11	
OxCAP-MH (n = 79)	Improved	25	<b>13 (52%)</b>	12 (48%)	0 (0%)	25	<b>13 (52%)</b>	12 (48%)	0 (0%)
	Stable	38	10 (26%)	<b>19 (50%)</b>	9 (24%)	38	9 (24%)	<b>25 (66%)</b>	4 (11%)
	Deteriorated	16	1 (6%)	10 (63%)	<b>5 (31%)</b>	16	2 (13%)	8 (50%)	<b>6 (38%)</b>
ICECAP-A (n = 88)	Improved	28	<b>15 (54%)</b>	12 (43%)	1 (4%)	28	<b>17 (61%)</b>	9 (32%)	2 (7%)
	Stable	46	10 (22%)	<b>28 (61%)</b>	8 (17%)	46	6 (13%)	<b>34 (74%)</b>	6 (13%)
	Deteriorated	14	1 (7%)	7 (50%)	<b>6 (43%)</b>	14	3 (21%)	8 (57%)	<b>3 (21%)</b>
EQ-5D-5L descriptive system (n = 94)	Improved	20	<b>11 (55%)</b>	7 (35%)	2 (10%)	20	<b>7 (35%)</b>	11 (55%)	2 (10%)
	Stable	62	16 (26%)	<b>33 (53%)</b>	13 (21%)	61	18 (30%)	<b>35 (57%)</b>	8 (13%)
	Deteriorated	18	1 (10%)	7 (70%)	<b>2 (20%)</b>	10	3 (30%)	6 (60%)	<b>1 (10%)</b>
EQ-5D VAS (n = 90)	Improved	26	<b>13 (50%)</b>	8 (31%)	5 (19%)	26	<b>13 (50%)</b>	9 (35%)	4 (15%)
	Stable	51	13 (25%)	<b>31 (61%)</b>	7 (14%)	51	14 (27%)	<b>32 (63%)</b>	5 (10%)
	Deteriorated	13	1 (8%)	8 (62%)	<b>4 (31%)</b>	12	0 (0%)	10 (83%)	<b>2 (17%)</b>

\*Changes in instrument scores between baseline and 9-month follow-up were categorised as improved, worsened or no change, definition of groups is based on the difference in SEM; values in agreement are in bold

Table 7 contrasts the number of patients who improved, remained stable and deteriorated based on the capability, HRQoL and the anchor instruments. All four measures identified the majority of patients in agreement with the anchor instruments, but the EQ-5D-5L descriptive system performed worst. Each instrument classified similar proportion of patients as “Stable” (50–74%), indicating similar sensitivity to change. There was no significant difference in terms of sensitivity to change between the investigated instruments.

## Agreement analysis

The Bland and Altman analysis showed that the OxCAP-MH and ICECAP-A having poorer agreement with EQ-5D-5L descriptive system than with each other or the EQ-5D VAS scale (Fig. 2). There is small average discrepancy between the four instruments; however, the limits of agreement were wider and therefore more ambiguous in the comparisons with EQ-5D-5L descriptive system than in the direct comparison of OxCAP-MH, ICECAP-A and EQ-5D VAS.

(Insert Fig. 2)

## Discussion

This paper aimed to contribute to the utilisation of the capability approach in mental health by empirically demonstrating that two instruments embedded in the capability framework but with a different approach to development show different psychometric properties when deployed on the same patient cohort. To our knowledge, this is the first paper to empirically compare the two most commonly used capability instruments in the area of mental health and compare them simultaneously to HRQoL, measured by the EQ-5D-5L descriptive system and the EQ-5D VAS. The study confirmed that both the OxCAP-MH and ICECAP-A instruments possess good psychometric properties among patients with severe mental health problems. In particular, this paper further confirmed the construct validity of both OxCAP-MH and ICECAP-A questionnaires. Both questionnaires are well correlated with self-reported measures of symptoms of anxiety (assessed with GAD) and general mental health wellbeing (e.g. WEMWBS); and relatively well correlated with instruments measuring depressive symptoms (assessed with BDI) and self-worth/self-esteem (assessed with RSES).

Different aspects of the analysis confirmed that the capability instruments had stronger associations and were more correlated to each other than to the HRQoL instruments, which implies that the capability instruments may be seen supplementary rather than complementary in their concept. The Bland Altman plots showed that the OxCAP-MH and ICECAP-A had poorer agreement with EQ-5D-5L than with each other. The results of the EFA of the items of both capability instruments and the EQ-5D-5L demonstrate that the capability instruments measure concepts beyond the standard interpretation of health because all items of the HRQoL measure loaded onto one factor, whilst the capability instruments spread across multiple factors. The results of the EFA suggest that the four factors represent different aspects of wellbeing measurement. Factor one could be linked to a narrower interpretation of health, but also including independence and suitable accommodation. Factor two includes items related to settlement and security aspects, where the negative loading of access to interesting activities indicates that this might be an auxiliary concept. Most of the ICECAP-A and OxCAP-MH items loaded on factor three, previously interpreted as internal psycho-social aspect of capabilities. These alternative loadings to factors two and three demonstrate the difference between the internal and external aspects of freedom within the capability approach. The findings of the current study are in line with a qualitative validation study of the Hungarian version of the OxCAP-MH, i.e. most domains in factor two and four of the EFA in this study are associated with the internal aspects of freedom, whilst factor three can be linked to external aspects (23). The two domains of OxCAP-MH related to the capabilities of appreciating people and nature loaded on a separate, fourth factor, indicating that this concept is supplementary and moves beyond the evaluative space included within the ICECAP-A or EQ-5D-5L

instruments. Nevertheless, some domains of both the OxCAP-MH and ICECAP-A load on the same factor together with the EQ-5D-5L items. This confirms the findings of Laszewska et al. (5); namely that the OxCAP-MH indeed measures some aspects of HRQoL and can be considered supplementary rather than complementary to EQ-5D-5L. An interesting finding of the current study is that the “Suitable accommodation” also loaded on to the common factor with EQ-5D-5L, suggesting an association with health. Previous studies exploring EFA between ICECAP-A and EQ-5D measures found that these two instruments share only a few common factors (13, 15). This study confirmed that there is a weak overlap between these two instruments, and only the “Being independent” domain of ICECAP-A loads on the same factor as the EQ-5D-5L items. These findings suggest that both capability instruments can be considered supplementary rather than complementary to EQ-5D-5L.

In contrast to most previous papers, this analysis presents relatively weak correlations between the OxCAP-MH, the ICECAP-A and the EQ-5D-5L descriptive system in the area of mental health. The OxCAP-MH was compared to the EQ-5D-3L and – 5L instruments in a mixed mental health population context and found correlation coefficients between 0.45–0.66 (5, 11). Similar correlations were observed between the ICECAP-A and the EQ-5D instruments when they were compared for opiate dependent patients. The study by Goranitis et al. found that ICECAP-A and EQ-5D-5L have similar construct validity when compared to other clinical measures (17). The slightly different results of the current study confirm previously identified weaknesses of the EQ-5D-5L instrument to measure HRQoL in severely ill mental health patients (52). Our results also confirm the findings of a study comparing ICECAP-A and EQ-5D-5L instruments in the area of depression, which concluded that instruments designed specifically to measure depression and mental health explained a greater proportion of the variation in ICECAP-A than the EQ-5D-5L (53).

In terms of sensitivity to change, no significant differences were observed between the two capability instruments, and the EQ-5D VAS performed better than the EQ-5D-5L descriptive system. The ICECAP-A seems to be slightly more correlated with generic measures, including the EQ-5D-5L descriptive system and the EQ-5D VAS, whilst the OxCAP-MH seems to be more highly correlated with disease-specific measures. This could be explained by either its supplementary nature, or the fact that the OxCAP-MH is a more detailed and longer questionnaire. The OxCAP-MH and ICECAP-A instruments are both embedded in the capability approach, but they were developed with a different approach, and this study has shown that they thereby show different psychometric properties when deployed on the same patient cohort. A major advantage of using the ICECAP-A in economic evaluations is the availability of its preference-based value set and its shorter length, which reduces the burden for respondents. Future research could explore the relationship of preference-based scores once those become available for the OxCAP-MH instrument if the relevant scale anchors allow.

Limitation of this research include a restricted number of data points compared to the number of items. Hence, the robustness of the EFA may be limited. In addition, the lack of an objective scale, which could indicate whether a patient has improved and which could be used as an absolute anchor in the calculations of external responsiveness statistics and Bland-Altman plots, could have potentially introduced some bias. External responsiveness could not be assessed by methods which require a gold-standard anchor, such as the Receiver Operating Characteristic (ROC) curve analysis. The reason for this is that none of the instruments in this study could be used as appropriate reference standards because they are all patient-reported measures (16). The responsiveness statistics also relied on a relatively small number of patients in most identified groups.

## Conclusion

The main conclusion of this study is that assessing outcomes in terms of capability for schizophrenic patients with depression provide more information than use of the NICE recommended QALY. Both the OxCAP-MH and the ICECAP-A are valid instruments to measure the impacts of mental health interventions within the capability framework. The EQ-5D-5L descriptive system showed less sensitivity to capture change and its evaluative space is also limited compared to both capability instruments. The two capability instruments were more convergent with each other than with any HRQoL measure confirming the hypothesised more similar underlying ‘capability’ construct. On the other hand, none of them proved superior to the other one in the current context. Instead, they seem to have different pros and cons. Establishing the psychometric properties of an instrument is a continuous process and further research should replicate this analysis on a higher number of patients and in other disease areas to strengthen these conclusions and explore potential psychometric differences related to diagnosis. Comparisons of OxCAP-MH and/or ICECAP-A with other capability (e.g. Achieved Capabilities Questionnaire for Community Mental Health (54)) or wellbeing (e.g. ReQoL (55)) instruments developed for the area of mental health would further contribute to our understanding of their measurement characteristics.

## Abbreviations

BDI: Beck Depression Inventory

EFA: Exploratory Factor Analysis

GAD: General Anxiety Disorder

HRQoL: Health-Related Quality of Life

ICECAP-A: ICEpop CAPability measure for Adults

NICE: National Institute for Health and Care Excellence

OxCAP-MH: Oxford CAPabilities questionnaire-Mental Health

QALYs: Quality-Adjusted Life Years

RSES: Rosenberg Self-Esteem Scale

SEM: Standard Error of Measurement

SRM: Standardised Response Mean

VAS: Visual Analogue Scale

WEMWBS: Warwick-Edinburgh Mental Well-being Scale

## Declarations

Ethics approval and consent to participate: The PoMeT trial received ethical approval from the Berkshire Research Ethics Committee (REC ref 13/SC/0634). All participants provided written consent.

Consent for publication: All authors consent

Availability of data and materials: Not applicable

Funding: The PoMeT trial was funded by the National Institute for Health Research (NIHR) under its Research for Patient Benefit (RfPB) Programme (Grant Reference Number PB-PG-0712-28021). Joanna Coast is supported by the Wellcome Trust [205384/Z/16/Z].

Author's contributions: TH and JS conceived of the presented idea and developed the conceptual framework of this research. JS provided the resources to this study. TH conducted the analysis with input on different aspects of the study from JC, AL, TS and JS. TH took the lead in writing the manuscript in close consultation with JC, AL, TS and JS. All authors provided critical feedback and helped shape the research, analysis and manuscript. All authors approved the final manuscript.

Acknowledgements: We would like to say thank you to all contributors and participants of the PoMeT trial. We would also like to thank Georg Heinze for his comments on the statistical analyses used in this paper and David Mott who discussed a previous version of this paper at Winter 2020 meeting of the Health Economics Study Group in Newcastle, UK.

Conflict of interest: JC has led the development of the ICECAP measures. JS has led the development of the OxCAP-MH measure. The remaining authors declare that they have no conflict of interest.

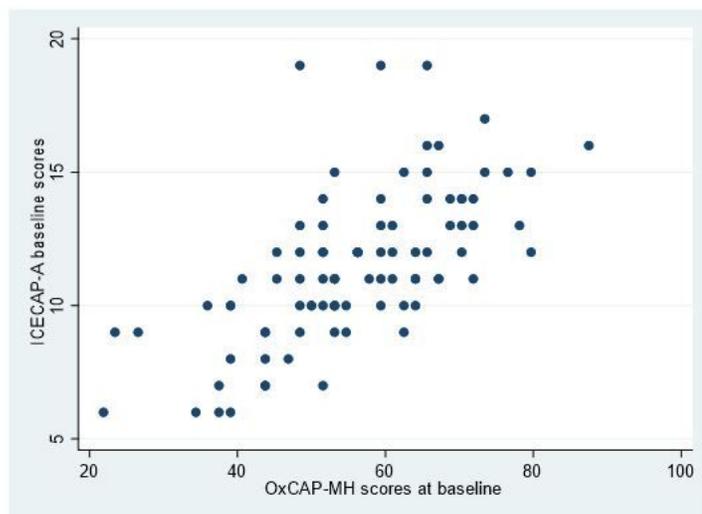
## References

1. Sen A. Commodities and capabilities. Amsterdam New York New York. for the U.S.A.: Elsevier Science Pub. Co.; N.Y., U.S.A: North-Holland Sole distributors; 1985.
2. Sen A. Development as freedom: Oxford University Press; 1999.
3. White RG, Imperiale MG, Perera E. The Capabilities Approach: Fostering contexts for enhancing mental health and wellbeing across the globe. *Globalization and Health*. 2016;12(1).
4. Helder TMCJ, Laszewska A, Stamm T, Simon J. Capability instruments in economic evaluations of health-related interventions – A comparative review of the literature. *Qual Life Res*. 2020;29:1433–64.
5. Łaszewska A, Schwab M, Leutner E, et al. Measuring broader wellbeing in mental health services: validity of the German language OxCAP-MH capability instrument. *Qual Life Res*. 2019.
6. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. *Br Med Bull*. 2010;96(1):5–21.
7. NICE  
NICE. Developing NICE guidelines: the manual 2014 (April 2017 update) - Incorporating economic evaluation. National Institute for Health and Clinical Excellence; 2017.
8. Nederland Z. Guideline for economic evaluations in healthcare 2016 [Available from: <https://english.zorginstituutnederland.nl/publications/reports/2016/06/16/guideline-for-economic-evaluations-in-healthcare>].
9. Francis JB. S. SCIE's approach to economic evaluation in social care. London: Social Care Institute for Excellence; 2011.
10. Simon JA, Gray P, Rugkasa A, Yeeles J, Burns K. T. Operationalising the capability approach for outcome measurement in mental health research. *Soc Sci Med*. 2013;98:187–96.
11. Vergunst F, Jenkinson C, Burns T, Anand P, Gray A, Rugkåsa J, et al. Psychometric validation of a multi-dimensional capability instrument for outcome measurement in mental health research (OxCAP-MH). *Health and Quality of Life Outcomes*. 2017;15(1).
12. Al-Janabi H, Peters TJ, Brazier J, Bryan S, Flynn TN, Clemens S, et al. An investigation of the construct validity of the ICECAP-A capability measure. *Quality of life research: an international journal of quality of life aspects of treatment care rehabilitation*. 2013;22(7):1831–40.
13. Davis JC, Liu-Ambrose T, Richardson CG, Bryan S. A comparison of the ICECAP-O with EQ-5D in a falls prevention clinical setting: are they complements or substitutes? *Qual Life Res*. 2013;22(5):969–77.
14. Keeley T, Al-Janabi H, Lorgelly P, Coast J. A qualitative assessment of the content validity of the ICECAP-A and EQ-5D-5L and their appropriateness for use in health research. *PLoS ONE*. 2013;8(12).

15. Keeley T, Coast J, Nicholls E, Foster NE, Jowett S, Al-Janabi H. An analysis of the complementarity of ICECAP-A and EQ-5D-3 L in an adult population of patients with knee pain. *Health and Quality of Life Outcomes*. 2016;14(1).
16. Goranitis I, Coast J, Al-Janabi H, Latthe P, Roberts TE. The validity and responsiveness of the ICECAP-A capability-well-being measure in women with irritative lower urinary tract symptoms. *Qual Life Res*. 2016;25(8):2063–75.
17. Goranitis I, Coast J, Day E, Copello A, Freemantle N, Seddon J, et al. Measuring Health and Broader Well-Being Benefits in the Context of Opiate Dependence: The Psychometric Performance of the ICECAP-A and the EQ-5D-5L. *Value in Health*. 2016;19(6):820–8.
18. Simon JMS, Laszewska A, Yeeles K, Burns T, Gray A. Cost and quality-of-life impacts of community treatment orders (CTOs) for patients with psychosis: Economic evaluation of the OCTET trial. *Social Psychiatry and Psychiatric Epidemiology*. 2020.
19. Al-Janabi HF, Coast TN. J. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Qual Life Res*. 2012;21(1):167–76.
20. Steel C, van der Gaag M, Korrelboom K, Simon J, Phiri P, Baksh MF, et al. A randomised controlled trial of positive memory training for the treatment of depression within schizophrenia. *BMC Psychiatry*. 2015;15(1).
21. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry*. 1961;4:561–71.
22. Simon J, Łaszewska A, Leutner E, Spiel G, Churchman D, Mayer S. Cultural and linguistic transferability of the multi-dimensional OxCAP-MH capability instrument for outcome measurement in mental health: The German language version. *BMC Psychiatry*. 2018;18(1).
23. Helter TMKI, Kanka A, Varga O, Kalman J, Simon J. Internal and external aspects of freedom in the application of the capability approach – the case study of developing a linguistically and culturally valid Hungarian version of the OxCAP-MH well-being questionnaire. unpublished manuscript.
24. Flynn TN, Huynh E, Peters TJ, Al-Janabi H, Clemens S, Moody A, et al. Scoring the Icecap-a Capability Instrument. Estimation of a UK General Population Tariff. *Health Econ*. 2015;24(3):258–69.
25. Chen G, Ratcliffe J, Kaambwa B, McCaffrey N, Richardson J. Empirical Comparison Between Capability and Two Health-Related Quality of Life Measures. *Soc Indic Res*. 2018;140(1):175–90.
26. Linton MJ, Mitchell PM, Al-Janabi H, Schlander M, Richardson J, Iezzi A, et al. Comparing the German Translation of the ICECAP-A Capability Wellbeing Measure to the Original English Version: Psychometric Properties across Healthy Samples and Seven Health Condition Groups. *Appl Res Qual Life*. 2018.
27. Khan MA, Richardson J. Variation in the apparent importance of health-related problems with the instrument used to measure patient welfare. *Qual Life Res*. 2018;27(11):2885–96.
28. Keeley T, Al-Janabi H, Nicholls E, Foster NE, Jowett S, Coast J. A longitudinal assessment of the responsiveness of the ICECAP-A in a randomised controlled trial of a knee pain intervention. *Qual Life Res*. 2015;24(10):2319–31.
29. Tang C, Xiong Y, Wu H, Xu J. Adaptation and assessments of the Chinese version of the ICECAP-A measurement. *Health and Quality of Life Outcomes*. 2018;16(1).
30. Birmingham Uo. ICECAP-A 2019 [Available from: <https://www.birmingham.ac.uk/research/activity/mds/projects/HaPS/HE/ICECAP/ICECAP-A/index.aspx>].
31. Engel L, Mortimer D, Bryan S, Lear SA, Whitehurst DGT. An Investigation of the Overlap Between the ICECAP-A and Five Preference-Based Health-Related Quality of Life Instruments. *PharmacoEconomics*. 2017;35(7):741–53.
32. Afentou N, Kinghorn P. A Systematic Review of the Feasibility and Psychometric Properties of the ICEpop CAPability Measure for Adults and Its Use So Far in Economic Evaluation. *Value in Health*. 2020;23(4):515–26.
33. Payakachat N, Ali MM, Tilford JM. Can The EQ-5D Detect Meaningful Change? A Systematic Review. *PharmacoEconomics*. 2015;33(11):1137–54.
34. EuroQol - a. new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16(3):199–208.
35. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–36.
36. Foundation ER. 2019 [Available from: <https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/valuation-standard-value-sets/>].
37. Association AP. Diagnostic and statistical manual of mental disorders. 4th ed text revision ed. Washington, DC: American Psychiatric Association; 2000.
38. Spitzer RL, Kroenke K, Williams JB, Lowe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med*. 2006;166(10):1092–7.
39. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Lowe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Ann Intern Med*. 2007;146(5):317–25.
40. Rosenberg M. Society and the adolescent self-image. Princeton: Princeton University Press; 1965.
41. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health Qual Life Outcomes*. 2007;5:63.
42. Streiner DL, Norman GR, Cairney J. Health measurement scales: A practical guide to their development and use. Oxford Oxford University Press; 2015.
43. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19(4):539–49.
44. Cohen J. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). 1988.
45. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol*. 2000;53(5):459–68.
46. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PMM. On assessing responsiveness of health-related quality of life instruments: Guidelines for instrument evaluation. *Qual Life Res*. 2003;12(4):349–62.

47. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* 2008;61(2):102–9.
48. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol.* 1999;52(9):861–73.
49. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care.* 1999;37(5):469–78.
50. Martin Bland J, Altman D. STATISTICAL METHODS FOR ASSESSING AGREEMENT, BETWEEN TWO METHODS OF CLINICAL MEASUREMENT. *The Lancet.* 1986;327(8476):307–10.
51. Watson PF, Petrie A. Method agreement analysis: A review of correct methodology. *Theriogenology.* 2010;73(9):1167–79.
52. Brazier J. Is the EQ–5D fit for purpose in mental health? *Br J Psychiatry.* 2010;197(5):348–9.
53. Mitchell PM, Al-Janabi H, Byford S, Kuyken W, Richardson J, Iezzi A, et al. Assessing the validity of the ICECAP-A capability measure for adults with depression. *BMC Psychiatry.* 2017;17(1).
54. Sacchetto B, Aguiar R, Vargas-Moniz MJ, Jorge-Monteiro MF, Neves MJ, Cruz MA, et al. The capabilities questionnaire for the community mental health context (CQ-CMH): A measure inspired by the capabilities approach and constructed through consumer-researcher collaboration. *Psychiatr Rehabil J.* 2016;39(1):55–61.
55. Keetharuth AD, Brazier J, Connell J, Bjorner JB, Carlton J, Taylor Buck E, et al. Recovering Quality of Life (ReQoL): a new generic self-reported outcome measure for use with people experiencing mental health difficulties. *The British journal of psychiatry: the journal of mental science.* 2018;212(1):42–9.
56. Kaiser HF. A second generation little jiffy. [journal article] *Psychometrika.* 1970;35(4):401–15. doi:10.1007/bf02291817.
57. Hutcheson GD, a. S. N. *The Multivariate Social Scientist: Introductory Statistics Using Generalized Linear Models.* Thousand Oaks: SAGE; 1999.
58. Bartlett MS. A note on multiplying factors for various chi-squared approximations. *Journal of the Royal Statistical Society Series B.* 1954;16:296–8.
59. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PMM. On assessing responsiveness of health-related quality of life instruments: Guidelines for instrument evaluation. [journal article]. *Qual Life Res.* 2003;12(4):349–62. doi:10.1023/a:1023499322593.
60. Shapiro A, ten Berge JMF. Statistical inference of minimum rank factor analysis. *Psychometrika.* 2002;67(1):79–94. doi:10.1007/bf02294710.
61. Lorenzo-Seva U. (2013). Why rotate my data using Promin? Technical report. U.R.I.V. Department of Psychology. Tarragona.
62. Lorenzo-Seva U. Promin: A Method for Oblique Factor Rotation. *Multivar Behav Res.* 1999;34(3):347–65. doi:10.1207/S15327906MBR3403\_3.
63. Sonntag M, Konnopka A, Leichsenring F, Salzer S, Beutel ME, Herpertz S, et al. Reliability, validity and responsiveness of the EQ-5D in assessing and valuing health status in patients with social phobia. [journal article]. *Health Quality of Life Outcomes.* 2013;11(1):215. doi:10.1186/1477-7525-11-215.

## Figures



**Figure 1**

Correlations between OxCAP-MH and ICEAP-A level sum scores at baseline

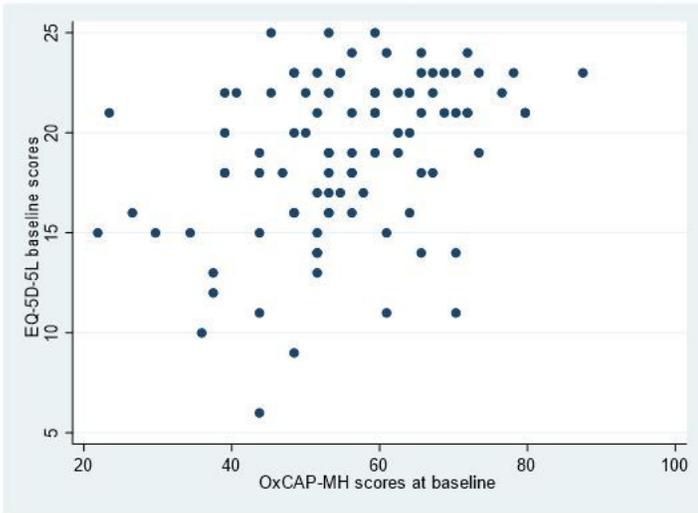


Figure 2

Correlations between OxCAP-MH and EQ-5D-5L descriptive system scores at baseline

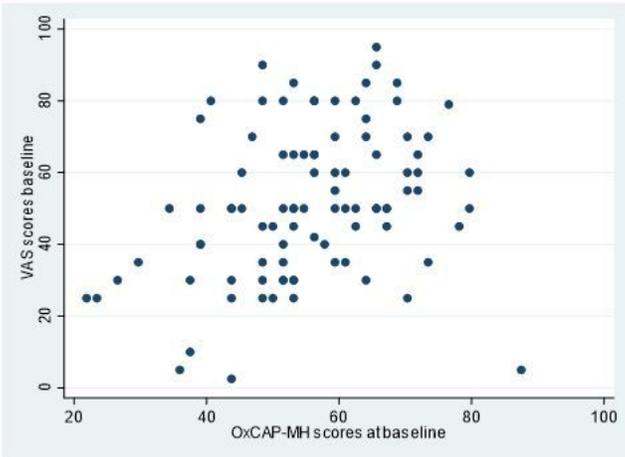


Figure 3

Correlations between OxCAP-MH and VAS scores at baseline

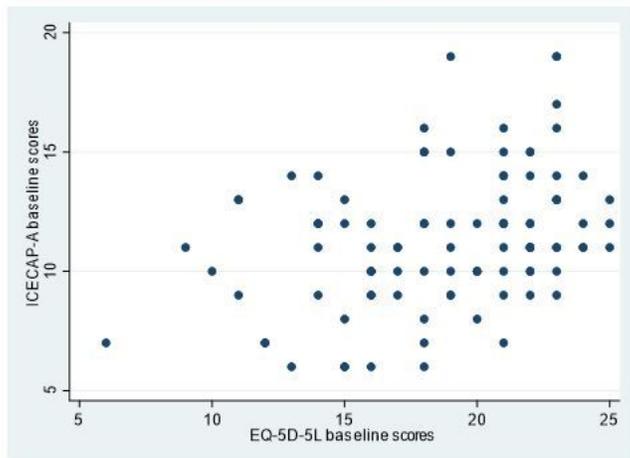


Figure 4

Correlations between ICEAP-A and EQ-5D-5L descriptive system level sum scores at baseline

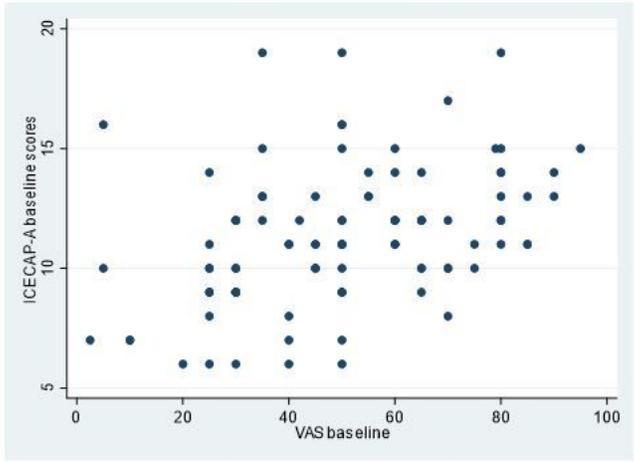


Figure 5

Correlations between ICECAP-A level sum scores and VAS scores at baseline

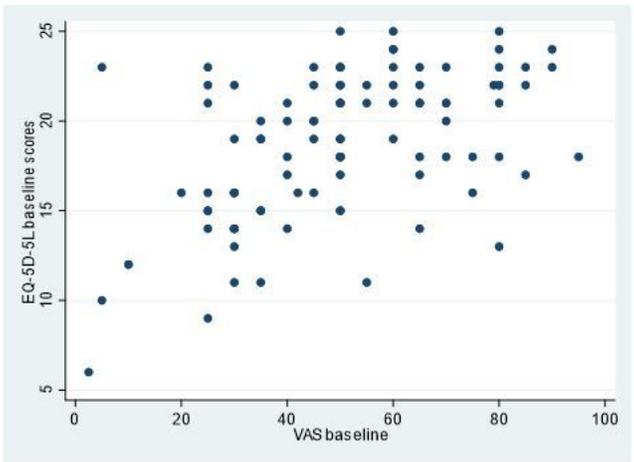


Figure 6

Correlations between EQ-5D-5L descriptive system level sum scores and VAS scores at baseline

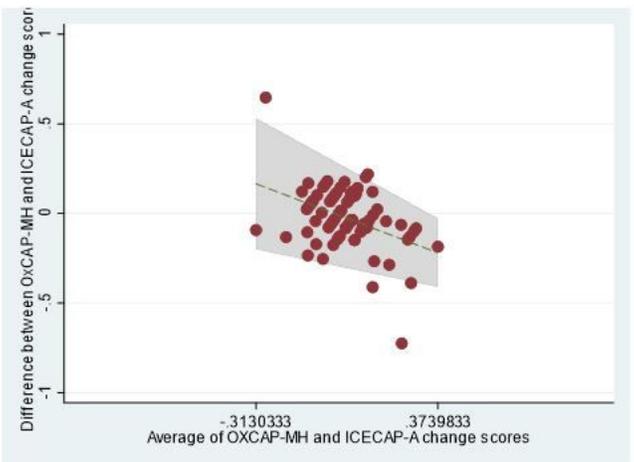


Figure 7

Bland-Altman plot of difference in OxCAP-MH vs ICECAP-A change scores

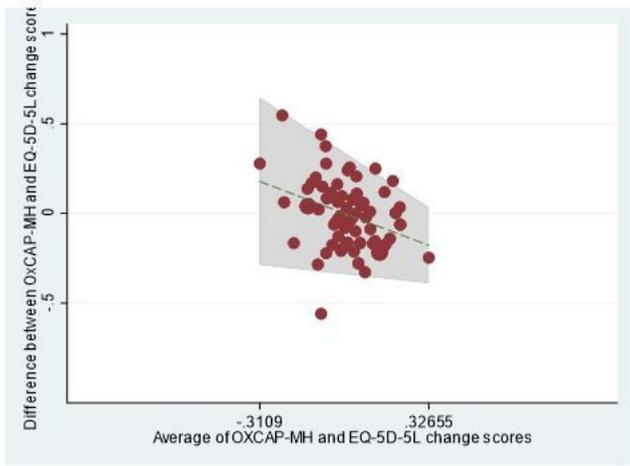


Figure 8

Bland-Altman plot of difference in OxCAP-MH vs EQ-5D-5L descriptive system change scores

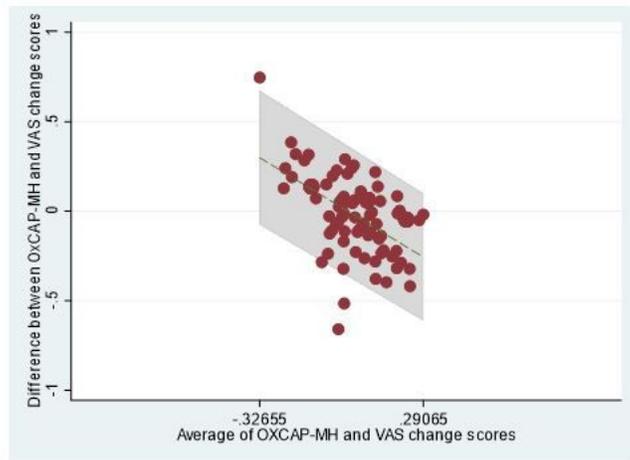


Figure 9

Bland-Altman plot of difference in OxCAP-MH vs VAS change scores

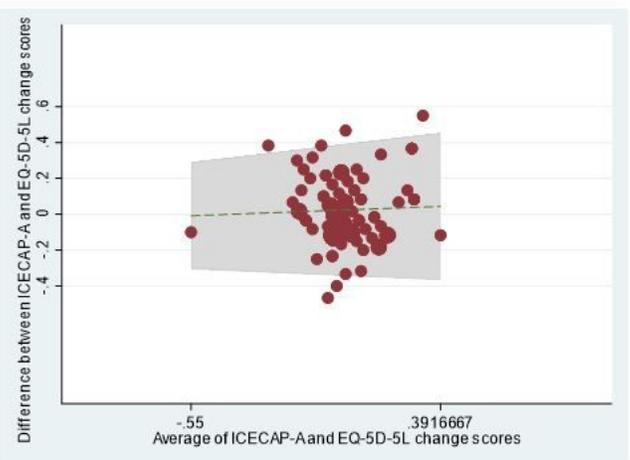


Figure 10

Bland-Altman plot of difference in ICECAP-A and EQ-5D-5L descriptive system change scores

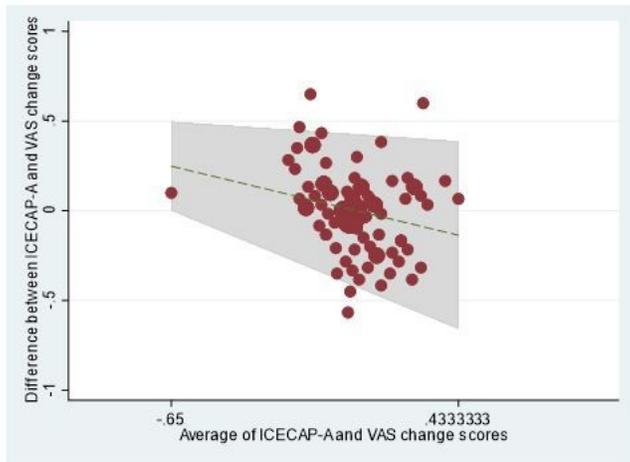


Figure 11

Bland-Altman plot of difference of ICECAP-A and VAS change scores

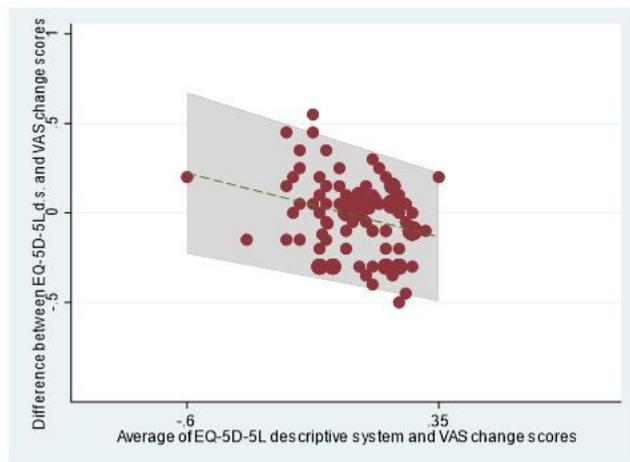


Figure 12

Bland-Altman plot of difference of EQ-5D-5L descriptive system and VAS change scores

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementarymaterialOxCAPMHICECAPAcomparison.pdf](#)