Assefa *et al. Health and Quality of Life Outcomes* (2025) 23:118
https://doi.org/10.1186/s12955-025-02444-1

Health and Quality of Life Outcomes

RESEARCH Open Access

Assessing measurement properties of EQ-5D-5L and ICECAP-A in patients with chronic depression in England

Esubalew Assefa[1], Victoria Bird[2], Stefan Priebe[3], Philip McNamee[4] and Yan Feng[1*]

Abstract

Background EQ-5D and ICECAP-A instruments are widely used as outcome measures in economic evaluation. This study aimed to evaluate the validity and responsiveness of EQ-5D-5L and ICECAP-A for economic evaluation in patients with chronic depression in England.

Methodology Data from 365 patients with chronic depression who took part in a cluster randomised controlled trial were used. The EQ-5D-5L, ICECAP-A and four additional outcome measures (MANSA, BDI-II, MADRS and CGI-S) were administrated at baseline and 12-months post-randomisation. We assessed the floor/ceiling effect, structural validity, convergent validity, known-group validity, and responsiveness of EQ-5D-5L and ICECAP-A in this population.

Results Exploratory factor analysis with pooled EQ-5D-5L and ICECAP-A items suggested a two-factor solution (i.e. psychosocial wellbeing and physical functioning factors). The EQ-5D-5L index and ICECAP-A index were positively correlated with each other ($\rho = 0.578$, $p < 0.001$). The two measures correlated the strongest with the BDI-II, and the weakest with CGI-S. Both EQ-5D-5L index and ICECAP-A index showed good discriminative ability. Mental health condition specific measures (BDI-II, MADRS and CGI-S) and a subjective quality of life measure (MANSA) reported better responsiveness than EQ-5D-5L index and ICECAP-A index. The ICECAP-A index reported better responsiveness than the EQ-5D-5L index. The responsiveness of the EQ-5D-5L was largely driven by the anxiety/depression dimension, followed by the usual activities dimension. For the ICECAP-A, four attributes of the instrument reported good responsiveness except for the autonomy attribute.

Conclusions EQ-5D-5L and ICECAP-A showed good validity and responsiveness among patients with chronic depression in England. We recommend ICECAP-A as an outcome measure in economic evaluation for medical interventions that aim to improve the psychosocial wellbeing of adults with chronic depression for the instrument's structural validity and its better responsiveness than the EQ-5D-5L.

Trial registration The trial was registered on 13/06/2019. The registration number is ISRCTN11301686.

Keywords EQ-5D-5L, ICECAP-A, Economic evaluation, Psychometrics, Depression, English NHS

*Correspondence:
Yan Feng
yan.feng@qmul.ac.uk
[1]Health Economics and Policy Research Unit, Centre for Evaluation and Methods, Wolfson Institute of Population Health, Queen Mary University of London, 58 Turner Street, London E1 2AB, UK
[2]Institute for Public Health and Wellbeing, University of Essex, Colchester, UK
[3]Centre for Psychosocial Medicine, University of Hamburg, Hamburg, Germany
[4]Unit for Social and Community Psychiatry, Centre for Psychiatry and Mental Health, Wolfson Institute of Population Health, Queen Mary University of London, London, UK

Introduction

Economic evaluation is a widely used method to assess the value for money of medical interventions [1]. The health benefits in economic evaluation are often expressed in quality-adjusted life years (QALYs). EQ-5D is one of the most widely used generic instruments to generate QALYs [2, 3]. Previous literature raised concerns about the appropriateness to use the EQ-5D instrument in patients with mental health disorders [4].

A recent systematic literature review on the measurement properties of the 5-level version of the EQ-5D instrument (EQ-5D-5L) [5] identified two studies [6, 7] on depression. These two studies limited their assessments of the EQ-5D-5L to convergent and known-groups validity. Furthermore, data in these two studies were collected from the general population, where patients with depression were identified by self-report rather than through clinical assessment. A more recent study assessed the validity and responsiveness of the EQ-5D-5L in patients with major depression in Spain [8]. Evidence on measurement properties of the EQ-5D-5L in this study was limited by its data collection strategy, i.e., only one mental health condition specific measure, Patient Health Questionnaire-9 (PHQ-9), was applied.

The ICEpop CAPability measure for Adults (ICECAP-A) is a generic measure of capability wellbeing for adults. It has been tested for its validity in adult depression population as an outcome measure. However, there is limited evidence on its responsiveness [9]. This literature gap was clearly identified in a recent systematic literature review which examined the evidence on the measurement properties of the ICECAP-A instrument and its use in economic evaluation [10]. The study highlighted the need for more evidence on the measurement properties of the ICECAP-A in patients with depression. A more recent publication compared the measurement properties between two capability measures (ICECAP-A and Oxford CAPabilities questionnaire-Mental Health (OxCAP-MH)) and two generic quality of life measures (EQ-5D-5L index and EuroQol Visual Analogue Scale (EQ-VAS)) in people with schizophrenia and depression in the UK [11]. The study concluded that ICECAP-A and OxCAP-MH provided more relevant, mental health-specific information than the EQ-5D-5L index or the EQ-VAS. However, the findings are limited by the small sample size (N = 100), which might affect the robustness of the exploratory factor analysis and the responsiveness analyses. Further research on the measurement properties of the ICECAP-A in patients with depression is clearly needed.

This study therefore aims to address the gaps identified in the literature by assessing the measurement properties of the EQ-5D-5L and ICECAP-A instruments in patients with chronic depression in England. We examined the validity and responsiveness of the two instruments using data from a cluster randomised controlled trial.

Methods

Trial background

Data were obtained from "Tackling Chronic Depression" (TACK) – a multi-site cluster randomised controlled trial (RCT) (ISRCTN11301686) [12]. The trial aimed to evaluate the effectiveness and cost-effectiveness of a technology-supported solution-focused intervention known as DIALOG + in the treatment of patients with chronic (recurrent) depression in England. Details of the trial are published elsewhere [13]. Briefly, TACK is a pragmatic multi-centre RCT conducted across nine NHS Trust sites. The recruitment began in July 2019 and ended in February 2022. Clinicians and their patients (who together form a cluster) were randomly allocated to either the treatment or active control group. The DIALOG + intervention includes two parts: (1) firstly patients are asked to self-rate their satisfaction with different life domains and treatment aspects using DIALOG scale and then (2) conduct structured discussion between clinicians and patients using a four-step approach based on the principles of solution-focused therapy [14]. Patients in the active control group received treatment as usual. In addition, control participants were asked to self-rate satisfaction on the 11 areas of the DIALOG scale at the end of each routine session. The study has been approved by the NHS Wales Research Ethics Committee 6 (19/WA/0160).

Participants

Eligible patients (N = 365) were between 18 and 100 years old, exhibited symptoms of depression or non-psychotic low mood with a duration of illness of at least 2 years, were receiving treatment from an NHS mental health service with regular contact with clinicians, had the capacity to provide informed consent and the ability to speak and understand English, have a score of less than 5 on Manchester Short Assessment of Quality of Life (MANSA) and a score of no less than 10 on Montgomery-Åsberg Depression Rating Scale (MADRS).

Outcome data collection

Patient-level data were collected at baseline, 3-, 6- and 12-months post-randomisation. As the trial interventions lasted for 12 months and not all outcomes were measured at all time points, our analyses focused on data collected at baseline (pre-intervention) and 12 months post-randomisation (after intervention). We analysed six outcome measures, namely EQ-5D-5L index, ICECAP-A index, MANSA score, Beck Depression Inventory – Second Edition (BDI-II score), MADRS score and Clinical Global Impression Scale (CGI-S) score.

EQ-5D

EQ-5D is a generic preference-based health-related quality of life measure. Data were collected using the 5L version of the instrument [15]. The EQ-5D-5L instrument captures individuals' health across five dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression). Each dimension has five severity levels (no problem, slight problem, moderate problem, severe problem and extreme problem). Following the National Institute for Health and Care Excellence (NICE) recommendations [3], we derived the 5L index from the 5L descriptive system data by applying for a mapping function (mapping from the UK EQ-5D-3L utilities which are in a range of -0.594 and 1) [16, 17].

ICECAP-A

ICECAP-A is a generic preference-based measure of capability wellbeing [18]. The measure comprises five attributes, i.e. stability (feeling settled and secure), attachment (love, friendship, and support), autonomy (being independent), achievement (achievement and progress), and enjoyment (enjoyment and pleasure). Each attribute has four response levels ranging from 1 (absence of capability) to 4 (full capability). ICECAP-A index was estimated by applying the UK tariff for the instrument [19]. The index values range from 0 (no capability) to 1 (full capability).

MANSA

MANSA [20] is a quality of life instrument which has been widely used with mental health patients. The subjective component of the instrument comprises 12 questions that measure individuals' satisfaction with their life as a whole, employment condition, financial situation, number and quality of friendships, leisure activities, accommodation, personal safety, living situation, sex life, relationship with family, physical health, and mental health. Participants were asked to self-rate on each item using a 7-point scale from 1 (couldn't be worse) to 7 (couldn't be better). A MANSA score is obtained by taking the average responses from the 12 questions, where higher scores indicate greater satisfaction. Trial participants had to have an average score of less than 5 to be eligible for inclusion in the trial.

BDI-II

BDI-II [21] is a 21-item self-reported inventory that assess the presence and severity of depression symptoms. Participants were asked to indicate how they have been feeling in the past two weeks on the 21 items. Each item corresponds to a symptom of depression on a 4-point scale ranging from 0 (not having the symptom) to 3 (having it all the time). A total BDI-II score is obtained by taking the sum of responses to the 21 items and ranges between 0 and 63.

MADRS

MADRS [22] is a 10-item observer-rated scale which was designed to assess severity of depression. The scale was completed through patient interviews with trained researchers. MADRS comprises 10 items including apparent sadness, reported sadness, inner tension, reduced sleep, reduced appetite, concentration difficulties, lassitude, inability to feel, pessimistic thoughts, and suicidal thoughts. Each item is scored on a range of 0 (normal) to 6 (severe). The total MADRS score ranges from 0 (no depression) to 60 (very severe depression). Trial participants had to have a MADRS score of no less than 10, at baseline, to be eligible for inclusion.

CGI-S

CGI-S [23] is a single item clinician rated scale. It asked clinicians to rate "*how mentally ill is the patient at this time given his/her total clinical experience with this particular population*" on a 7-point scale. The scale is in a range of 1 (normal, not at all ill) to 7 (amongst the most severely ill patients).

Analyses

Floor and ceiling effects

We examined the floor and ceiling effects of the six outcome measures using baseline data. The percentages of participants who scored the worst or best possible values on each outcome measure were calculated. Floor and ceiling effects were considered present if 15% or more of participants achieved the lowest or highest possible values respectively [24, 25].

Structural validity

We performed exploratory factor analysis to assess the underlying structure of the EQ-5D-5L and ICECAP-A instruments. By analysing the ten pooled items from the two instruments (i.e. 5 items from the EQ-5D-5L and 5 items from ICECAP-A), exploratory factor analysis aimed to identify a smaller number of latent variables or underlying factors that can explain the variance amongst the ten items [11, 26]. We applied the Kaiser–Meyer–Olkin (KMO) test to check if data were suitable for factor analysis (i.e. KMO value below 0.50 as unacceptable, 0.50–0.59 as miserable, 0.60–0.69 as mediocre, 0.70–0.79 as middling, 0.80–0.89 as meritorious, and 0.90 and above as marvellous) [27]. We also performed the Bartlett's test of sphericity to explore if correlation matrix among variables is significantly different from identity matrix. A significant test result, with p-value less than 0.05, was considered as an indication that the data is appropriate for factor analysis [28]. The number

of factors were determined by the eigenvalue of greater than or equal to 1. We performed the factor analysis with polychoric correlation to account for the ordinal nature of the variables, followed by promax rotation of loading matrix. To simplify the presentation, we reported factor loadings larger than 0.3. We also reported the correlation between factors.

Convergent validity

We assessed the convergent validity of the EQ-5D-5L index and ICECAP-A index by examining how they correlate with each other, and with MANSA, BDI-II, MADRS and CGI-S scores. We hypothesised that the EQ-5D-5L index and ICECAP-A index positively correlate with each other, and with MANSA scores. We also hypothesised that the EQ-5D-5L index and ICECAP-A index negatively correlate with BDI-II, MADRS and CGI-S scores. We reported the Spearman's correlation coefficients using baseline and the 12 months follow-up data. To reflect the ordinal nature of CGI-S score, we also reported the polyserial correlation coefficients between CGI-S and the EQ-5D-5L index and ICECAP-A index at baseline and the 12 months follow-up.

Known-groups validity

We assessed the ability of EQ-5D-5L index and ICECAP-A index to discriminate between known-groups of participants based on their severity of depression. We used BDI-II and MADRS scores and their respective cut-off points to define the severity of depression among participants. BDI-II score of 0–13 is considered as no depression, 14–19 as mild, 20–28 as moderate, and 29–63 as severe depression [21]. For MADRS score, 0–6 is considered as normal/no depression, 7–19 as mild depression, 20–34 as moderate depression, and 35–60 as severe depression [29]. For each outcome measure, we calculated Cohen's d standardised absolute effect size (AES), i.e. the difference in mean scores between two adjacent severity groups divided by the standard deviation of scores for the milder of the two groups. We also conducted the Kruskal-Wallis test on each outcome measure to check whether scores from different severity groups come from the same distribution.

Responsiveness

Responsiveness refers to the ability of an instrument to detect changes over time [30]. We assessed the responsiveness of the EQ-5D-5L index and ICECAP-A index in two ways, i.e. the measure's ability to detect (1) changes in general over time (without anchor) and (2) clinically meaningful changes over time (with anchor) [31]. We used longitudinal data from participants in the intervention arm at baseline (pre-intervention) and the 12 months follow-up (post-intervention).

For approach (1), we assessed the responsiveness of the EQ-5D-5L index, ICECAP-A index, five EQ-5D-5L dimensions, and five ICECAP-A attributes using the Standardised Response Mean (SRM). The SRM is calculated by dividing the mean change of an outcome measure from baseline to the 12 months follow-up by the standard deviation of the change. Previous literature suggested that an SRM greater than 0.8 is considered as large responsiveness, 0.5 to 0.8 as moderate responsiveness, 0.2 to 0.5 as small responsiveness, and less than 0.2 as negligible [32]. For comparison, we also assessed the responsiveness for MANSA, BDI-II, MADRS and CGI-S.

For approach (2), we assessed the responsiveness of the EQ-5D-5L index, ICECAP-A index, five EQ-5D-5L dimensions, and five ICECAP-A attributes using two depression symptom-based measures (BDI-II and MADRS) as external anchors. We first divided participants in the intervention arm into improved, deteriorated, and unchanged groups according to their changes in health status (measured by BDI-II and MADRS) from baseline to the 12 months follow-up. For the BDI-II, a six-point score change was used to define the reliable change [33]. We applied a two-point score change as the reliable change for MADRS [34]. We then assessed the responsiveness for EQ-5D-5L index, ICECAP-A index, five EQ-5D-5L dimensions, and five ICECAP-A attributes by the SRM of the improved, deteriorated and unchanged groups.

This study followed the "COSMIN Reporting Guideline 2.0 for studies on measurement properties of patient-reported outcome measures" on reporting structural validity, hypotheses testing for construct validity (i.e. convergent and known-groups validity) and responsiveness [35]. All statistical analyses were performed using Stata version 18 [36].

Results

Participant characteristics and baseline outcomes

In total, there were 365 participants randomised into the trial with 51.8% in the treatment arm and 48.2% in the active control arm (Table 1). The average age of participants was 46.7 years (SD = 15.17). The majority were female (65.3%), White British (87.6%), not married (70.4%), and had completed further or higher education (62.2%). Around half of the participants were unemployed (47.1%). On average, participants reported having received depression treatment for 13.8 years. Participants self-reported poor quality of life, poor capability, moderate to severe depression, and were rated as moderately to markedly ill by their clinicians at baseline.

Almost all participants completed EQ-5D-5L ($N = 360$ completed with 1.4% missing) and ICECAP-A ($N = 357$ completed with 2.2% missing) at baseline. At 12 months follow-up, the proportion of missingness increased for all

Table 1 Baseline characteristics and outcome of participants

Variable names	Overall *N* = 365	Intervention arm *N* = 189	Active control arm *N* = 176
Age in years, mean (SD)	46.7 (15.2)	45.9 (15.1)	47.6 (15.3)
Sex, n (% female)	237 (65.3)	125 (66.5)	112 (64.0)
Marital Status, n (%)			
Married	108 (29.6)	48 (25.4)	60 (34.1)
Not married	257 (70.4)	141 (74.6)	116 (65.9)
Ethnicity, n (%)			
White British	319 (87.6)	163 (86.2)	156 (89.1)
Other	45 (12.4)	26 (13.8)	19 (10.9)
Educational attainment, n (%)			
Primary or less, or secondary	129 (35.3)	68 (36.0)	61 (34.7)
Tertiary, higher education	227 (62.2)	117 (61.9)	110 (62.5)
Other general education	9 (2.5)	4 (2.1)	5 (2.8)
Employment status, n (%)			
Employed	89 (24.4)	41 (21.7)	48 (27.3)
Unemployed	172 (47.1)	94 (49.7)	78 (44.3)
Other	104 (28.5)	54 (28.6)	50 (28.4)
Treatment years for depression	13.8	13.9	13.7
Outcome measures, mean (SD)			
EQ-5D-5L index	0.395 (0.295)	0.381 (0.300)	0.410 (0.291)
ICECAP-A index	0.481 (0.179)	0.476 (0.177)	0.486 (0.182)
MANSA	3.538 (0.753)	3.521 (0.708)	3.555 (0.800)
BDI-II	36.367 (11.664)	36.532 (12.024)	36.186 (11.293)
MADRS	30.507 (8.662)	31.188 (8.604)	29.783 (8.689)
CGI-S	4.230 (0.912)	4.169 (0.870)	4.299 (0.954)

Table 2 Factor structure of the EQ-5D-5L and ICECAP-A

Dimension/attribute	Factor 1 (psychosocial wellbeing)	Factor 2 (physical functioning)	Uniqueness
EQ-5D-5L			
Mobility		**0.864**	0.315
Self-care	-0.314	**0.522**	0.472
Usual activities	**-0.548**		0.475
Pain/discomfort		**0.779**	0.429
Anxiety/depression	**-0.716**		0.498
ICECAP-A			
Stability	**0.666**		0.603
Attachment	**0.424**		0.840
Autonomy	**0.346**	-0.307	0.684
Achievement	**0.622**		0.617
Enjoyment	**0.696**		0.568
Factor correlation	-0.48		

Factor loadings below 0.30 are not shown in the table. Bold numbers in the table refer to the larger factor loadings between the two factors for a given instrument and dimension/attribute

six outcome measures. Additional details on missing data are reported in Supplementary File Table S1. Complete case analysis was applied to handle missing values.

Floor and ceiling effects

There was no evidence of floor or ceiling effects for any of the outcome measures. One participant (0.3%) reported the highest possible score on CGI-S.

Structural validity

The results from KMO (= 0.81) and Bartlett's ($\chi^2(45) = 912.23$, $p < 0.001$) tests suggested that the data is appropriate for factor analysis. From the exploratory factor analysis, two-factor solution was optimal with eigenvalue greater than one. The two factors moderately correlated with each other (= 0.48) (Table 2). Factor 1 had meaningful loadings on all ICECAP-A attributes and two dimensions of the EQ-5D-5L (usual activities and

Table 3 Correlation coefficients between EQ-5D-5L index, ICECAP-A index and four other measures at baseline and 12 months follow-up

	(a) Baseline		(b) 12 months	
	EQ-5D-5L index	**ICECAP-A index**	**EQ-5D-5L index**	**ICECAP-A index**
ICECAP-A index [a]	0.578*		0.567*	
MANSA [a]	0.487*	0.589*	0.537*	0.725*
BDI-II [a]	-0.583*	-0.615*	-0.658*	-0.755*
MADRS [a]	-0.506*	-0.492*	-0.694*	-0.741*
CGI-S [a]	-0.216*	-0.254*	-0.345*	-0.445*
CGI-S [b]	-0.207*	-0.284*	-0.399*	-0.453*

[a] Spearman's correlation coefficient; [b] Polyserial correlation coefficient; * $p < 0.001$

anxiety/depression). Factor 2 had meaningful loadings on three dimensions of the EQ-5D-5L (mobility, self-care, and pain/discomfort). We labelled factors 1 and 2 as *psychosocial wellbeing* and *physical functioning* respectively.

Convergent validity

All correlation coefficients at baseline showed the expected signs (block (a) in Table 3). The EQ-5D-5L index and ICECAP-A index positively correlated with each other ($\rho = 0.578$, $p < 0.001$), and with MANSA. The two outcome measures negatively correlated with BDI-II, MADRS and CGI-S scores. The EQ-5D-5L index and ICECAP-A index correlated the strongest with the BDI-II, and the weakest with CGI-S. At 12 months follow-up, we observed stronger correlation between the outcome measures than at baseline, except for the correlation relationship between the EQ-5D-5L index and ICECAP-A index (block (b) in Table 3). The polyserial correlation coefficients between CGI-S and the two outcome measures (i.e. the EQ-5D-5L index and ICECAP-A index) at baseline and 12 months are reported in the last row of Table 3. The results are consistent with that from the Spearman's correlation test.

Known-groups validity

The mean value of the EQ-5D-5L index and ICECAP-A index by the level of depression severity, defined by the BDI-II and MADRS scores, are presented in Table 4. Under both definitions of severity levels, the EQ-5D-5L index and ICECAP-A index decreased with an increase in depression severity. The AES are in a range of 0.615 and 1.046. The results from Kruskal-Wallis test suggested that the medians are statistically significantly different between depression severity groups (defined by the BDI-II and MADRS scores) for the EQ-5D-5L index and ICECAP-A index.

Responsiveness

The results from analysing the six outcome measures, using data from baseline to the 12 months follow-up, for their responsiveness are reported in Table 5. All measures showed improvements at the 12 months follow-up. The two generic outcome measures (i.e. EQ-5D-5L index and ICECAP-A index) are less responsive compared to MANSA and the three condition specific measures (i.e. BDI-II, MADRS and CGI-S). The SRM is the smallest for the EQ-5D-5L index (0.246), followed by ICECAP-A index (0.488). For MANSA and the three mental health specific measures, the absolute SRM is reported in a range of 0.569 and 0.887. The driving force behind the responsiveness of EQ-5D-5L index are improvements in the anxiety/depression dimension (SRM = 0.264) and usual activities dimension (SRM = 0.167). The responsiveness of the other three dimensions of the EQ-5D-5L are negligible (SRM from 0.041 to 0.082). For ICECAP-A index, four attributes report SRM in a range of 0.286 and 0.389, while the autonomy attribute contributes little to the responsiveness (SRM = 0.071).

Table 4 Known-groups validity check using severity level defined by BDI-II and MADRS

Depression levels	***N***	**EQ-5D-5L index [d]**	**AES [e]**	**ICECAP-A index**	**AES**
BDI-II [a]					
Minimal/Mild	26	0.679 (0.141)		0.701 (0.152)	
Moderate	61	0.546 (0.240)	0.615	0.565 (0.156)	0.883
Severe	240	0.332 (0.283)	0.780	0.442 (0.158)	0.779
Total	327	0.400 (0.290)		0.485 (0.176)	
p-value [b]		< 0.001		< 0.001	
MADRS[c]					
Mild	43	0.622 (0.212)		0.666 (0.165)	
Moderate	196	0.457 (0.267)	0.641	0.506 (0.149)	1.046
Severe	122	0.214 (0.268)	0.911	0.370 (0.158)	0.890
Total	361	0.394 (0.295)		0.480 (0.180)	
p-value [b]		< 0.001		< 0.001	

[a] Minimal/Mild (Minimal = 0–13, Mild = 14–19), Moderate = 20–28, Severe = 29–63; [b] Kruskal-Wallis test; [c] Mild depression = 7–19, Moderate depression = 20–34, Severe depression = 35–60; [d] Mean (SD); [e] AES (Absolute Effect Size): Cohen's d standardised absolute effect size comparing a group with the adjacent less severe group

Table 5 Responsiveness based on score change by outcome measure

Outcome measures	Time points	*N*	Mean	SD	SRM
EQ-5D-5L index	Baseline	185	0.381	0.300	
	12 months	133	0.462	0.329	
	Δ (12 months – baseline)	132	0.065	0.264	0.246
Mobility	Baseline	187	1.957	1.154	
	12 months	135	2.007	1.136	
	Δ (12 months – baseline)	134	0.030	0.735	0.041
Self-care	Baseline	188	1.984	1.082	
	12 months	135	1.830	0.981	
	Δ (12 months – baseline)	134	-0.082	1.004	-0.082
Usual activities	Baseline	188	2.840	1.200	
	12 months	135	2.570	1.231	
	Δ (12 months – baseline)	134	-0.216	1.294	-0.167
Pain/discomfort	Baseline	187	2.497	1.220	
	12 months	135	2.467	1.239	
	Δ (12 months – baseline)	134	-0.045	1.075	-0.042
Anxiety/Depression	Baseline	188	3.670	1.012	
	12 months	134	3.231	1.156	
	Δ (12 months – baseline)	133	-0.331	1.254	-0.264
ICECAP-A index	Baseline	185	0.476	0.177	
	12 months	131	0.581	0.216	
	Δ (12 months – baseline)	127	0.092	0.189	0.488
Stability	Baseline	188	1.793	0.791	
	12 months	135	2.185	0.839	
	Δ (12 months – baseline)	134	0.358	0.921	0.389
Attachment	Baseline	188	2.441	0.796	
	12 months	134	2.694	0.852	
	Δ (12 months – baseline)	133	0.248	0.848	0.293
Autonomy	Baseline	185	2.589	0.941	
	12 months	135	2.726	0.868	
	Δ (12 months – baseline)	131	0.061	0.857	0.071
Achievement	Baseline	188	1.904	0.663	
	12 months	132	2.152	0.756	
	Δ (12 months – baseline)	131	0.229	0.800	0.286
Enjoyment	Baseline	188	2.005	0.580	
	12 months	135	2.296	0.829	
	Δ (12 months – baseline)	134	0.261	0.803	0.325
MANSA	Baseline	189	3.521	0.708	
	12 months	141	4.243	0.899	
	Δ (12 months – baseline)	141	0.670	0.807	0.831
BDI-II	Baseline	171	36.532	12.024	
	12 months	119	26.933	14.370	
	Δ (12 months – baseline)	111	-9.721	12.196	-0.797
MADRS	Baseline	186	31.188	8.604	
	12 months	135	24.681	12.549	
	Δ (12 months – baseline)	133	-5.820	10.230	-0.569
CGI-S	Baseline	160	4.169	0.870	
	12 months	97	3.041	1.154	
	Δ (12 months – baseline)	85	-1.200	1.352	-0.887

SRM = Standardised Response Mean

The results from anchor-based approach are reported in Table 6. ICECAP-A index reported better responsiveness than the EQ-5D-5L index for the deteriorated and improved groups using the BDI-II score change as the external anchor. Using the MADRS score change as the external anchor, again, we observed ICECAP-A index responded better than the EQ-5D-5L index for the improved group. While the EQ-5D-5L index performed better in capturing responsiveness from the deteriorated group than the ICECAP-A index. The dimension/attribute level analyses for the two instruments are reported in Tables S2 and S3 in Supplementary File. The results suggested that the key dimension for the EQ-5D-5L instrument to capture improvement at 12 months follow-up is the anxiety/depression dimension, followed by the usual activities dimension. For ICECAP-A, four attributes of the instrument contribute to detect the improvement at 12 months follow-up, while contribution from the autonomy attribute is negligible. These findings are consistent between applying BDI-II or MADRS as the external anchor.

Discussion

Summary of findings

In this study, we evaluated the validity and responsiveness of the EQ-5D-5L and ICECAP-A instruments in patients with chronic depression in England. We used data from a cluster RCT. First, the exploratory factor analysis suggested that the two-factor solution was the optimal approach, using pooled items from the EQ-5D-5L and ICECAP-A instruments, with all five attributes of the ICECAP-A and two dimensions of the EQ-5D-5L loaded to the "psychosocial wellbeing" factor, and other three dimensions of the EQ-5D-5L loaded to the "physical functioning" factor. Second, we found moderate correlation between the EQ-5D-5L index and ICECAP-A index, and with subjective quality of life (MANSA) and mental health condition specific measures (BDI-II, MADRS). The EQ-5D-5L index and ICECAP-A index weakly correlated with the clinician rating (CGI-S). Furthermore, our results suggested that the correlation between measures are generally stronger at 12 months follow-up than at baseline. Third, the EQ-5D-5L index and ICECAP-A index were able to discriminate between different severities of depression. The former discriminated the more severe level better, while the latter performed better for the milder depression level. Four, subjective quality of life (MANSA) and mental health condition specific measures (BDI-II, MADRS and CGI-S) showed better responsiveness than the EQ-5D-5L index and ICECAP-A index. Between the two generic outcome measures, ICECAP-A index reported better responsiveness than the EQ-5D-5L index. The responsiveness of the EQ-5D-5L at 12 months follow-up was largely driven by the anxiety/depression dimension followed by the usual activities dimension. The contribution from other three dimensions are negligible. For ICECAP-A, four attributes of the instrument (except for autonomy) well contributed to its responsiveness at 12 months follow-up.

Comparing to previous literature

The TACK sample, in terms of demographics and illness severity, was broadly representative of patients with chronic or recurrent depression in the UK. When looking at clinically diagnosed populations within UK primary care services, a large study recruiting from 42 GP practices across the UK showed that the average age of those with chronic or recurrent depression was 48 years, 75% were female, and the majority (96.3%) were White [37]. This is comparable to the age, gender and ethnicity reported in the TACK trial sample (i.e. average age was 46.7 years, 65.3% were female, and 87.6% were White). In terms of BDI-II scores, this study reported a lower average total score from their study sample (recruited from primary care) in comparison to the TACK trial sample (recruited from secondary care) (i.e. 32.52 vs. 36.37). This indicates a slightly higher depression severity in the TACK trial sample. Similar findings were reported in the FINDER study [38], where 608 depressed patients in the UK were recruited from a primary care setting. The study sample had an average age of 42.8 years, 61.2% were female, and average duration of depression was 5.3 years. Further, in a large descriptive epidemiological study of individuals in the UK BioBank, who were assessed for a lifetime history of mood disorder, demographics for (probable) recurrent major depression were similar to those presented in the TACK trial [39]. Mean age for recurrent major depression (moderate) was 55.4 years, percentage of females was 68.8%, and White ethnicity was 94.7%.

However, TACK participants self-reported poor health status compared to the general population. The average value of the EQ-5D-5L index (0.4) and ICECAP-A index (0.48) were much lower than the UK/English general population (i.e. EQ-5D-5L index = 0.876 [40], ICECAP-A index = 0.81 [41]).

The EQ-5D-5L index and ICECAP-A index did not show ceiling or floor effects in our study, although ceiling effect was previously reported by EQ-5D-5L instrument in depression patients [42].

Two studies [26, 43] explored the level of overlap between the EQ-5D-3L and ICECAP instruments using exploratory factor analysis. Both studies concluded that two-factor solution was the optimal. For ICECAP, four attributes strongly loaded to the same factor, while the other attribute (i.e., autonomy attribute in ICECAP-A or control attribute in ICECAP-O) loaded moderately to both factors. For EQ-5D-3L, anxiety/depression

Table 6 Responsiveness of two outcome measures using BDI-II score change and MADRS score change as external anchor

Outcome measures	Time points	Deteriorated group				Unchanged group				Improved group				
		N	Mean	SD	SRM	*N*	Mean	SD	SRM	*N*	Mean	SD	SRM	*p*-value [c]
BDI-II [a]														
EQ-5D-5L	Baseline	10	0.386	0.333		31	0.427	0.311		70	0.395	0.287		0.841
	12 months	10	0.306	0.408		30	0.334	0.328		70	0.542	0.296		0.005
	Δ (12 months – baseline)	10	-0.080	0.273	-0.294	30	-0.085	0.181	-0.467	70	0.147	0.252	0.584	< 0.001
ICECAP-A	Baseline	10	0.525	0.111		30	0.496	0.195		69	0.498	0.171		0.894
	12 months	10	0.448	0.154		30	0.486	0.197		69	0.657	0.205		< 0.001
	Δ (12 months – baseline)	10	-0.076	0.099	-0.775	29	-0.005	0.152	-0.035	68	0.163	0.182	0.899	< 0.001
MADRS[b]														
EQ-5D-5L	Baseline	27	0.314	0.316		17	0.312	0.352		88	0.450	0.280		0.063
	12 months	24	0.221	0.304		17	0.257	0.347		88	0.570	0.278		< 0.001
	Δ (12 months – baseline)	24	-0.084	0.218	-0.385	17	-0.055	0.236	-0.231	88	0.120	0.260	0.462	< 0.001
ICECAP-A	Baseline	26	0.441	0.202		17	0.447	0.180		87	0.514	0.165		0.050
	12 months	24	0.418	0.176		15	0.532	0.168		88	0.637	0.201		< 0.001
	Δ (12 months – baseline)	23	-0.015	0.161	-0.091	15	0.073	0.169	0.430	86	0.126	0.189	0.669	0.003

[a] A six-point score change was used as a measure of reliable change. Deteriorated group refers to patients with a BDI-II score increase (i.e., worsening symptoms) of six or more. Improved group refers to patients with a BDI-II score decrease of six or more. And unchanged group refers to patients with a BDI-II score increase or decrease less than six

[b] A two-point score change was used as a measure of reliable change. Deteriorated group refers to patients with a MADRS score increase (i.e., worsening symptoms) of two or more. Improved group refers to patients with a MADRS score decrease of two or more. And unchanged group refers to patients with a MADRS score increase or decrease less than two

[c] Kruskal-Wallis test

dimension loaded strongly to one factor, and the remaining four dimensions loaded to the other factor. These findings are consistent with ours, except for in our study, the "usual activities" dimension of the EQ-5D-5L strongly loaded to the same factor as the anxiety/depression dimension. Both dimensions fall under the psychosocial wellbeing factor which captures individuals' social functioning (usual activities) and mental functioning (anxiety/depression) [2].

A recent study compared a few capability and health-related quality of life instruments in patients with schizophrenia and depression in the UK [11]. In comparison to their findings, our results showed stronger correlation between the EQ-5D-5L index and ICECAP-A index using data collected from chronic depression patients in England. Furthermore, our findings suggested that the correlation between generic quality of life and capability measures (EQ-5D-5L index and ICECAP-A index) and four other measures (MANSA, BDI-II, MADRS, CGI-S) are stronger at 12 months follow-up than at baseline. This is consistent with findings from previous literature that the EQ-5D-5L has stronger correlation with patient reported depression outcome measures for milder health states [44, 45].

Previous studies suggested that generic outcome measures are less responsive to health benefit changes than condition specific measures [46]. Our responsiveness analyses provide evidence to support this argument. We found that MANSA and the three mental health condition-specific outcome measures (i.e., BDI-II, MADRS, CGI-S) reported better responsiveness at the 12 months follow-up than the EQ-5D-5L index and ICECAP-A index. Furthermore, our results suggest that the ICECAP-A index was more responsive than the EQ-5D-5L index. Results from the dimension/attribute level analyses of the EQ-5D-5L and ICECAP-A instruments provided explanations for this finding. To be specific, we found that the responsiveness of the EQ-5D-5L and ICECAP-A instruments was largely driven by the dimensions/attributes under the *psychosocial wellbeing* factor (i.e., two out of five dimensions of the EQ-5D-5L instrument, and four out of five attributes of the ICECAP-A instrument) but not those under the *physical functioning* factor.

Strengths and limitations

The TACK trial dataset provides a unique opportunity to assess the measurement properties of two widely used generic preference-based outcome measures. The trial dataset is large (N = 365) despite the well-known difficulties in recruiting participants into depression trials. There is a general consensus that depression trials commonly experience particular challenges with recruitment and retention. Many studies fail to recruit their target sample of participants due to a number of reasons, including the affective symptoms, gatekeeping, and stigma [47]. The trial sample used in this study, although being recruited exclusively from secondary mental health care services, appeared to be similar to the chronic depression samples reported in primary care studies and broader epidemiological surveys. This indicates that findings are broadly generalisable to broader chronic depression populations. In addition to the EQ-5D-5L and ICECAP-A data, the trial also collected data using two depression symptom-based measures, one subjective quality of life measure, and a clinician rated scale. Those data enabled our validity and responsiveness assessments.

We report a few limitations of this study. First, we observed the proportion of missingness increased for all six outcome measures at 12 months follow-up. However, the TACK trial sample size is larger than the recommended rules of thumb for the analyses that we conducted [48, 49]. Second, although the Kaiser criterion is commonly used in psychometric studies and guides to determine the optimal number of factors in exploratory factor [50], the method has been criticized for its potential overestimation [51, 52]. Third, in the responsiveness analyses, we applied the widely used cut-off point for each of the two external anchors [21, 29]. However, there is no consensus on the value of reliable change for BDI-II and MADRS scores. Fourth, our trial sample does not represent the full depression patient population in England. The trial focused on patients with chronic depression, defined as with symptoms for more than 2 years. Over one third of patients who experienced an acute episode of depression in the UK developed chronic depression [13].

Implications

Our finding suggested that the EQ-5D-5L and ICECAP-A instruments have different underlying structures. The ICECAP-A instrument focuses on the psychosocial wellbeing of individuals, while the EQ-5D-5L instrument captures individuals' psychosocial wellbeing *and* physical functioning. In TACK, the DIALOG + intervention aimed to help clinicians to make their routine meetings with chronic depression patients therapeutically effective. The intervention is expected to improve quality of life and/or depression symptoms of the trial participants. By analysing the overall indices of the EQ-5D-5L and ICECAP-A, we showed that the ICECAP-A had better responsiveness than EQ-5D-5L at the 12 months follow-up. The dimension/attribute level analyses suggested that both instruments have good responsiveness from dimensions/attributes that the DIALOG + intervention targeted for. To be specific, for the EQ-5D-5L, anxiety/depression dimension showed good responsiveness at the 12 months follow-up and reported the strongest responsiveness across the five dimensions. The instrument's three

dimensions under physical functioning reported poor responsiveness. For the ICECAP-A, all five attributes focus on capturing individuals' psychosocial wellbeing. The attribute level analyses showed good responsiveness from four attributes at the 12 months follow-up. Based on our findings, we recommend ICECAP-A as an outcome measure in economic evaluation for medical interventions that aim to improve the psychosocial wellbeing of adults with chronic depression because of the instrument's structural validity and its better responsiveness than the EQ-5D-5L.

Conclusion

EQ-5D-5L and ICECAP-A instruments reported good validity and responsiveness among chronic depression patients in England. Because of the structural validity of ICECAP-A and its better responsiveness than the EQ-5D-5L, we recommend ICECAP-A as an outcome measure in economic evaluation for medical interventions that aim to improve the psychosocial wellbeing of adults with chronic depression.

Abbreviations

AES	Absolute Effect Size
BDI-II	Beck Depression Inventory – Second Edition
CGI-S	Clinical Global Impression Scale
EQ-5D-5L	5-level version of the EQ-5D instrument
EQ-VAS	EuroQol Visual Analogue Scale
ICECAP-A	ICEpop CAPability measure for Adults
KMO	Kaiser–Meyer–Olkin
MANSA	Manchester Short Assessment of Quality of Life
MADRS	Montgomery-Åsberg Depression Rating Scale
NICE	National Institute for Health and Care Excellence
OxCAP-MH	Oxford CAPabilities questionnaire-Mental Health
PHQ-9	Patient Health Questionnaire-9
QALYs	Quality-adjusted Life Years
RCT	Randomised Controlled Trial
SRM	Standardised Response Mean

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12955-025-02444-1.

Supplementary Material 1

Acknowledgements

We would like to thank Philip Kinghorn and participants at the Health Economists' Study Group summer 2024 Conference (Warwick, UK) and participants at the annual International Society for Quality of Life Research 2024 Conference (Cologne, Germany) for their helpful feedback and comments. TACK trial from which the data for analysis were obtained was funded by the National Institute for Health and Care Research (NIHR) Programme Grants for Applied Research (RP-PG-0615-200010).

Author contributions

YF conceptualised the study and developed the analysis plan. EA and YF conducted data analyses, drafted the first version of the manuscript, performed editing and formatting for all versions. VB, SP and PM provided expert clinical oversight (include interpreting the clinical data) and commented on previous versions of the manuscript. EA, VB, SP, PM, and YF approved the final manuscript. YF is the guarantor.

Funding

The TACK trial from which the data for analysis were obtained was funded by the National Institute for Health and Care Research (NIHR) Programme Grants for Applied Research (RP-PG-0615-200010). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Data availability

The data that support the findings of this study are available from the TACK trial team but restrictions apply to the availability of these data and are not publicly available.

Declarations

Ethics approval and consent to participate

Ethics approval for the study was granted by the NHS Wales Research Ethics Committee 6 (19/WA/0160). Informed consent was obtained from all individual participants included in the study.

Consent for publication

Informed consent was obtained from all individual participants included in the study.

Competing interests

YF is a member of the EuroQol Research Foundation, i.e. the copyright holder of the EQ-5D instruments. EA, VB, SP, and PM declare no competing interests.

Received: 7 January 2025 / Accepted: 8 October 2025
Published online: 17 November 2025

References

1. Brazier J. Measuring and valuing mental health for use in economic evaluation. J Health Serv Res Policy. 2008;13:70–5.
2. Brooks R. EuroQol: the current state of play. Health Policy. 1996;37:53–72.
3. NICE. NICE health technology evaluations: the manual [Internet]. 2022. Available from: https://www.nice.org.uk/process/pmg36/resources/nice-health-technology-evaluations-the-manual-pdf-72286779244741
4. Brazier J. Is the EQ–5D fit for purpose in mental health? Br J Psychiatry. 2010;197:348–9.
5. Feng Y-S, Kohlmann T, Janssen MF, Buchholz I. Psychometric properties of the EQ-5D-5L: a systematic review of the literature. Qual Life Res. 2021;30:647–73.
6. Mihalopoulos C, Chen G, Iezzi A, Khan MA, Richardson J. Assessing outcomes for cost-utility analysis in depression: comparison of five multi-attribute utility instruments with two depression-specific outcome measures. Br J Psychiatry. 2014;205:390–7.
7. Engel L, Chen G, Richardson J, Mihalopoulos C. The impact of depression on health-related quality of life and wellbeing: identifying important dimensions and assessing their inclusion in multi-attribute utility instruments. Qual Life Res. 2018;27:2873–84.
8. Bilbao A, Martín-Fernández J, García-Pérez L, Mendezona JI, Arrasate M, Candela R, et al. Psychometric properties of the EQ-5D-5L in patients with major depression: factor analysis and Rasch analysis. J Ment Health [Internet]. 2022 [cited 2024 May 27]; Available from: https://www.tandfonline.com/doi/abs/10.1080/09638237.2021.1875422
9. 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;17:1–13.
10. 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 Health. 2020;23:515–26.
11. Helter TM, Coast J, Łaszewska A, Stamm T, Simon J. Comparison of capability and health-related quality of life instruments in capturing aspects of mental well-being in people with schizophrenia and depression. BJPsych Open. 2022;8:e117.
12. ISRCTN. ISRCTN11301686: Tackling chronic depression (TACK II) [Internet]. [cited 2024 May 20]. Available from: https://www.isrctn.com/ISRCTN11301686
13. McNamee P, Matanov A, Jerome L, Kerry S, Walker N, Feng Y, et al. Clinical- and cost-effectiveness of a technology-supported and solution-focused

intervention (DIALOG+) in treatment of patients with chronic depression—study protocol for a multi-site, cluster randomised controlled trial [TACK]. Trials. 2022;23:1–13.
14. Priebe S, Kelley L, Omer S, Golden E, Walsh S, Khanom H, et al. The effectiveness of a Patient-Centred assessment with a Solution-Focused approach (DIALOG+) for patients with psychosis: A pragmatic Cluster-Randomised controlled trial in community care. Psychother Psychosom. 2015;84:304–13.
15. Herdman M, Gudex C, Lloyd A, Janssen MF, 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;20:1727–36.
16. Hernández Alava M, Pudney S. Equation 5Dmap: A command for mapping between EQ-5D-3L and EQ-5D-5L. Stata J. 2018;18:395–415.
17. Hernández Alava M, Pudney S, Wailoo A. Estimating the relationship between EQ-5D-5L and EQ-5D-3L: results from a UK population study. PharmacoEconomics. 2023;41:199–207.
18. Al-Janabi H, Flynn N, Coast T. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. Qual Life Res. 2012;21:167–76.
19. 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;24:258–69.
20. Priebe S, Huxley P, Knight S, Evans S. Application and results of the Manchester short assessment of quality of life (Mansa). Int J Soc Psychiatry. 1999;45:7–12.
21. Beck AT, Steer RA, Brown GK. BDI-II, Beck Depression Inventory: manual. Psychological Corporation; 1996.
22. Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry. 1979;134:382–9.
23. Guy W. ECDEU assessment manual for psychopharmacology. U.S. Department of Health, Education, and Welfare. Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration. National Institute of Mental Health, Psychopharmacology Research Branch, Division of Extramural Research Programs; 1976.
24. McHorney CA, Tarlov AR. Individual-Patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res. 1995;4:293–307.
25. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60:34–42.
26. Keeley T, Coast J, Nicholls E, Foster NE, Jowett S, Al-Janabi H. An analysis of the complementarity of ICECAP-A and EQ-5D-3L in an adult population of patients with knee pain. Health Qual Life Outcomes. 2016;14:1–5.
27. Kaiser HF. An index of factorial simplicity. Psychometrika. 1974;39:31–6.
28. Bartlett MS. The effect of standardization on a χ2 approximation in factor analysis. Biometrika. 1951;38:337–44.
29. Canadian Agency for Drugs and Technologies in Health. Validity of outcome measures. Aripiprazole (Abilify): Depression, Major Depressive Disorder (MDD) [Internet]. 2016 [cited 2024 Mar 18]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK409740/
30. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63:737–45.
31. 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;12:349–62.
32. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. J Clin Epidemiol. 2000;53:459–68.
33. Swan J, Sorrell E, MacVicar B, Durham R, Matthews K. Coping with depression: an open study of the efficacy of a group psychoeducational intervention in chronic, treatment-refractory depression. J Affect Disord. 2004;82:125–9.
34. Montgomery S. Clinically relevant outcomes measures in major depressive disorder. Int J Neuropsychopharmacol. 2008;11:1–338.
35. Gagnier JJ, de Arruda GT, Terwee CB, Mokkink LB, Elsman EBM, Firth AD, et al. COSMIN reporting guideline for studies on measurement properties of patient–reported outcome measures: version 2.0. Qual Life Res [Internet]. 2025; Available from: https://doi.org/10.1007/s11136-025-03950-x
36. StataCorp. Stata statistical software: release 18. College station. TX: StataCorp LLC.; 2023.
37. McMahon EM, Buszewicz M, Griffin M, Beecham J, Bonin E-M, Rost F, et al. Chronic and recurrent depression in primary care: Socio-Demographic Features, Morbidity, and costs. Int J Family Med. 2012;2012:316409.
38. Lenox-Smith A, Macdonald MTB, Reed C, Tylee A, Peveler R, Quail D, et al. Quality of life in depressed patients in UK primary care: the FINDER study. Neurol Ther. 2013;2:25–42.
39. Smith DJ, Nicholl BI, Cullen B, Martin D, Ul-Haq Z, Evans J, et al. Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: Cross-Sectional study of 172,751 participants. PLoS ONE. 2013;8:e75362.
40. Feng Y, Devlin N, Herdman M. Assessing the health of the general population in england: how do the three- and five-level versions of EQ-5D compare? Health Qual Life Outcomes. 2015;13:1–16.
41. Mitchell PM, Morton RL, Hiligsmann M, Husbands S, Coast J. Estimating loss in capability wellbeing in the first year of the COVID-19 pandemic: a cross-sectional study of the general adult population in the UK, Australia and the Netherlands. Eur J Health Econ. 2023;24:609–19.
42. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res. 2013;22:1717–27.
43. 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;22:969–77.
44. Sapin C, Fantino B, Nowicki M-L, Kind P. Usefulness of EQ-5D in assessing health status in primary care patients with major depressive disorder. Health Qual Life Outcomes. 2004;2:1–8.
45. Mann R, Gilbody S, Richards D. Putting the 'Q' in depression qalys: a comparison of utility measurement using EQ-5D and SF-6D health related quality of life measures. Soc Psychiat Epidemiol. 2009;44:569–78.
46. Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. J Clin Epidemiol. 2003;56:52–60.
47. Hughes-Morley A, Young B, Waheed W, Small N, Bower P. Factors affecting recruitment into depression trials: systematic review, meta-synthesis and conceptual framework. J Affect Disord. 2015;172:274–90.
48. Van Wilson CR, Morgan BL. Understanding power and rules of thumb for determining sample sizes. TQMP. 2007;3:43–50.
49. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD. What is sufficient evidence for the reliability and validity of Patient-Reported outcome measures? Value Health. 2007;10:S94–105.
50. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine: A Practical Guide [Internet]. Cambridge, UK: Cambridge University Press; 2011 [cited 2025 May 7]. Available from: http://ebookcentral.proquest.com/lib/gmul-ebooks/detail.action?docID=802925
51. Horn JL. A rationale and test for the number of factors in factor analysis. Psychometrika. 1965;30:179–85.
52. Hayton JC, Allen DG, Scarpello V. Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. Organizational Res Methods. 2004;7:191–205.

Publisher's note