CrossMark

ORIGINAL RESEARCH ARTICLE

An Investigation of the Overlap Between the ICECAP-A and Five Preference-Based Health-Related Quality of Life Instruments

Lidia Engel[1,2] · Duncan Mortimer[3] · Stirling Bryan[2,4] · Scott A. Lear[1,5] · David G. T. Whitehurst[1,2,6]

Abstract

Background The ICEpop CAPability measure for Adults (ICECAP-A) is a measure of capability wellbeing developed for use in economic evaluations. It was designed to overcome perceived limitations associated with existing preference-based instruments, where the explicit focus on health-related aspects of quality of life may result in the failure to capture fully the broader benefits of interventions and treatments that go beyond health. The aim of this study was to investigate the extent to which preference-based health-related quality of life (HRQoL) instruments are able to capture aspects of capability wellbeing, as measured by the ICECAP-A.

Methods Using data from the Multi Instrument Comparison project, pairwise exploratory factor analyses were conducted to compare the ICECAP-A with five preference-based HRQoL instruments [15D, Assessment of Quality of Life 8-dimension (AQoL-8D), EQ-5D-5L, Health Utilities Index Mark 3 (HUI-3), and SF-6D].

Results Data from 6756 individuals were used in the analyses. The ICECAP-A provides information above that garnered from most commonly used preference-based HRQoL instruments. The exception was the AQoL-8D; more common factors were identified between the ICECAP-A and AQoL-8D compared with the other pairwise analyses.

Conclusion Further investigations are needed to explore the extent and potential implications of 'double counting' when applying the ICECAP-A alongside health-related preference-based instruments.

✉ David G. T. Whitehurst
  david_whitehurst@sfu.ca

1 Faculty of Health Sciences, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada

2 Centre for Clinical Epidemiology and Evaluation, Vancouver Coastal Health Research Institute, Vancouver, BC, Canada

3 Faculty of Business and Economics, Centre for Health Economics, Monash University, Clayton, VIC, Australia

4 School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

5 Division of Cardiology, Providence Health Care, Vancouver, BC, Canada

6 International Collaboration on Repair Discoveries, University of British Columbia, Vancouver, BC, Canada

---

Key Points for Decision Makers

There is growing interest in measuring outcomes for economic evaluation in a way that goes beyond health-related quality of life (HRQoL). Operationalizing this work requires consideration of the overlap between these different approaches.

Variation in the dimensions of HRQoL included in current preference-based HRQoL instruments means that overlap with wellbeing and/or capability instruments [here, the ICEpop CAPability measure for Adults (ICECAP-A)] will differ.

In this study, the ICECAP-A provided additional complementary information when compared with the 15D, EQ-5D-5L, Health Utilities Index Mark 3 (HUI-3), and SF-6D, while there was substantial overlap between the ICECAP-A and Assessment of Quality of Life 8-dimension (AQoL-8D).

△ Adis

1 Introduction

Economic evaluation has become an important tool in many countries to inform decision makers about the value of alternative courses of action [1]. These evaluations usually take the form of a cost-utility analysis, where the outcome is measured in quality-adjusted life years (QALYs) [2]. The QALY has become the gold standard measure of health outcome in economic evaluation and is recommended by numerous health technology assessment agencies to assist in the allocation of scarce healthcare resources [3–5]. Indeed, an objective to 'maximize health' within a healthcare system is often operationalized by maximizing the number of QALYs gained from a fixed budget.

The focus on QALYs for resource allocation decisions in healthcare has been challenged for decades [6–8], with recent contributions drawing attention to areas such as public health [9], social care [10], mental health [11], and end-of-life care [12]. It is often argued that there are important benefits that cannot be measured in terms of health alone and that the evaluative space of economic evaluations should be more encompassing, allowing for the inclusion of broader benefits, such as wellbeing [13]. The term 'wellbeing' has been used inconsistently in the literature [14], although a distinction can be made between psychological wellbeing (a eudaimonic measure, i.e., a measure of flourishing such as self-acceptance or autonomy) [15] and subjective wellbeing (a hedonic measure, i.e., a measure of happiness and satisfaction) [16]. Another conceptualization of wellbeing has been offered by Amartya Sen, referred to as the capability approach [17], which distinguishes between capabilities (a person's opportunities to achieve wellbeing) and achieved functionings (the actual outcomes realized by individuals) [18]. The capability approach accounts for the fact that a person's capabilities (what a person *can* do) may differ from their functionings (what a person actually *does*) [19]. There is growing interest in Sen's capability approach within health economics, and for outcome measurements in economic evaluations in particular [20]. Recent efforts to operationalize the capability approach have led to the development of preference-based instruments for the measurement of capability wellbeing, suitable for use in economic evaluation. Three such measures have resulted from the Investigating Choice Experiments for the Preferences of Older People (ICEPOP) project: the ICEpop CAPability measure for Older Adults (ICECAP-O) [21], Adults (ICECAP-A) [19], and individuals at the end of life [ICECAP Supportive Care Measure (ICECAP-SCM)] [22].

Changes in guidelines for health technology assessments have recognised the potential importance of broader benefits in economic evaluation and have made provision for the measurement of capability wellbeing. For example, the National Institute for Health and Care Excellence in the UK has recommended the use of capability measures in economic evaluations for interventions that are associated with non-health benefits [23], yet little guidance has been provided in terms of what constitutes a health benefit or a non-health benefit, and which decision rules should be applied if using ICECAP instruments alongside other preference-based health-related quality of life (HRQoL) instruments. Dutch guidelines also advocate the use of ICECAP instruments for long-term care, where the focus of interventions might be more on improving a person's wellbeing rather than their health [5]. Because ICECAP instruments do not have 'QALY properties' (i.e., current values are not anchored onto the 'full health' to 'dead' scale but on a 'full capability' to 'no capability' scale [24]), the reference cases described in the UK and Dutch guidelines recommend supplementing cost-utility analysis (using the EQ-5D [25, 26]) with a cost-consequences analysis or cost-effectiveness analysis using an ICECAP instrument. The underlying intention is to capture explicitly broader aspects of capability wellbeing alongside health benefits.

In practice, decision makers may find it difficult to interpret and reconcile findings from such primary and supplementary analyses without further information describing the extent of overlap between measures of HRQoL and capability wellbeing. The extent of overlap between the ICECAP instruments and the three-level EQ-5D (EQ-5D-3L) has been examined in two previous studies. Davis and colleagues performed an exploratory factor analysis (EFA) comparing the ICECAP-O with the EQ-5D-3L in seniors enrolled in a falls prevention clinic [27], showing that the two instruments tapped into distinct and complementary factors. These results were confirmed by a second EFA, which compared the ICECAP-A with the EQ-5D-3L in an adult population of patients with knee pain [28].

Further research is needed to explore whether the same relationship holds in other clinical and non-clinical settings, as well as for other preference-based HRQoL instruments. Preference-based HRQoL instruments differ greatly in their coverage of physical, mental, and social health domains [29–32], as well as the extent to which 'non-health' items are included in the respective descriptive systems. These issues raise the potential for different degrees of overlap between preference-based HRQoL instruments (i.e., preference-based instruments that define health states) and measures of capability wellbeing. Such investigations are particularly important to avoid double counting when using HRQoL and capability wellbeing instruments simultaneously in health economic evaluations. In this context, double counting, where the same underlying concept of benefit is measured twice, could

occur explicitly (i.e., summing health and non-health benefits into a single metric) or implicitly (e.g., misguided interpretation of outcomes data from a cost-consequences analysis). The objectives of this work are to investigate the extent to which five preference-based HRQoL instruments capture aspects of capability wellbeing, as measured by the ICECAP-A, and to consider the implications of our findings within the context of other literature regarding capability wellbeing and economic evaluation.

2 Methods

2.1 Data Source

Data were obtained from the Multi Instrument Comparison (MIC) project, a multinational survey funded by Australia's National Health and Medical Research Council. Comprehensive details regarding the background, rationale, and administration of the MIC survey have been reported elsewhere [33]. Briefly, the aim of the MIC project was to compare several quality of life and wellbeing instruments across seven disease areas (in addition to a 'disease free' population) in six countries: Australia, Canada, Germany, Norway, UK, and USA. The MIC survey was administered online between February 2012 and May 2012 by a global survey company, CINT Pty Ltd.

2.2 Instruments

The MIC survey contained a comprehensive set of questions and standardized instruments [33]. In addition to questions about demographics, self-reported illnesses, and subjective wellbeing, all participants were asked to complete the ICECAP-A (with the exception of participants in Norway) and seven preference-based HRQoL instruments: 15D [34], Assessment of Quality of Life 4-dimension (AQoL-4D) [35], Assessment of Quality of Life 8-dimension (AQoL-8D) [36], EQ-5D-5L [26], Health Utilities Index Mark 3 (HUI-3) [37], Quality of Well-Being Scale Self-Administered (QWB-SA) [38], and SF-6D (based on the 36-item Short Form health survey version 2 (SF-36v2) [39]) [40]. Instruments were administered in a randomized order to account for order-effect bias [41].

For the analyses reported in the current paper, a decision was made to focus on the 35-item AQoL-8D (rather than the 12-item AQoL-4D) because of the more comprehensive descriptive system and the greater potential for overlap. The QWB-SA was also excluded because the measurement scale used for many items provides nominal data. These data would require transformation to meet the requirements for the statistical analysis performed, and such transformations render the analysis meaningless because the

descriptive system has been modified. An overview of the dimensions and items contained within the preference-based HRQoL instruments included in this analysis (15D, AQoL-8D, EQ-5D-5L, HUI-3, and SF-6D) is provided in Online Supplementary Material I, with more comprehensive details available elsewhere [30]. The ICECAP-A comprises five dimensions (lay descriptions used by the instrument developers are included in brackets): stability (an ability to feel settled and secure), attachment (an ability to have love, friendship, and support), autonomy (an ability to be independent), achievement (an ability to achieve and progress in life), and enjoyment (an ability to experience enjoyment and pleasure). Each dimension comprises one question with four levels of response, ranging from full capability to no capability [19].

2.3 Statistical Analysis

Exploratory factor analyses were conducted in Mplus 7.4 (Muthén & Muthén, Los Angeles) [42]. In all pairwise comparisons (i.e., item-level responses for the ICECAP-A compared with item-level responses for each of the other instruments, namely 15D, AQoL-8D, EQ-5D-5L, HUI-3, and SF-6D), EFA was used to ascertain the number of unique underlying latent factors that were associated with the items covered by the respective preference-based HRQoL instrument and the ICECAP-A. The purpose of the EFA was to explore the underlying structure for a set of measures and to determine whether or not the ICECAP-A instrument measures something unique, i.e., a construct or constructs not captured by current preference-based HRQoL instruments. Output from EFA includes factor loadings, which reflect the strength and direction of association between each item and each of the common factors. Higher factor loadings indicate that more of the variance in the observed variables (i.e., items from the descriptive systems of the instruments being compared) is attributable to the latent variable (i.e., the common factor) [43]. The axes of the initial factor analysis were rotated using the *geomin* oblique rotation. Oblique rotation permits correlations between common factors, which is to be expected when all items measure aspects of a person's quality of life. Pearson correlation coefficients were used to examine the extent of the relationship between factors (factors are considered as continuous variables); correlations were interpreted as weak (0.10–0.30), moderate (0.30–0.50), or strong (>0.50) [44]. Weighted least-square means and variance adjusted model estimation were applied to account for the ordinal nature of the item-level data.

The factor model and the number of common factors for each pairwise analysis were selected using the following procedure. The first step comprised an examination of

eigenvalues. Eigenvalues are numerical values that correspond to the variance in the items accounted for by each of the common factors [43]. More specifically, an eigenvalue is the sum of the squared factor loadings for a given factor. Model selection based on eigenvalues typically entails comparison of eigenvalues against the Kaiser criterion, where the number of factors with eigenvalues >1 gives the number of common factors to be specified in the model [43]. Evaluation against the Kaiser criterion was supplemented with inspection of scree plots, which are graphical representations of the eigenvalues plotted in a descending order. Model selection based on scree plots typically involves identification of the last substantial drop in the magnitude of the eigenvalues and retention of common factors prior to this drop [45].

Scree plots also guided the identification of increases (decreases) in the number of factors suggested by the Kaiser criterion that return large gains (small losses) in the variance. Three model fit indices were used to further quantify such gains (losses). The root mean square error of approximation (RMSEA) estimates goodness of fit as the discrepancy between the model and the data per degree of freedom for the model [45]; RMSEA values were interpreted as indicating a close (<0.05), acceptable (0.05–0.08), marginal (0.081–0.1), or poor (>0.1) fit [43]. The Tucker–Lewis Index (TLI) and Comparative Fit Index (CFI) were also used. These goodness-of-fit estimates indicate how much better a model fits the data compared with a baseline model that assumes no relationship exists between any of the variables [46]. For both the TLI and CFI, values >0.9 indicate a 'good' model fit [47].

Using more than one criterion to guide the selection of the number of factors raises the possibility of seemingly conflicting results (e.g., a situation where the Kaiser criterion suggests a two-factor model, whereas model fit statistics suggest a three-factor model). Within EFA, it is important to recognize that the objective is not to arrive at the 'true' or 'correct' number of factors but to estimate the patterns of correlations among observed variables and to simplify the data so that these patterns of correlations can be more easily interpreted [43]. Where selection based on the Kaiser criterion, scree plot, and model fit did not yield a 'clean' factor structure, models with an increased number of factors were explored to see whether this improved the interpretation of the model (i.e., the interpretability of each set of items in the respective factors). A clean factor structure is given when item loadings are all >0.3 on at least one factor, and there are no or few cross-factor loadings (i.e., items that load >0.3 on more than one factor) [48]. Where expansion of the number of factors failed to remove cross-loadings, the parsimonious model with fewer factors suggested by the Kaiser criterion, scree plot, and model fit statistics was selected as the preferred model.

Once a preferred factor model was identified for each pairwise comparison, using the procedure described above, overlap between the ICECAP-A and the respective HRQoL instrument was examined using the following criteria: (1) the number of common factors shared by both instruments, and (2) the extent to which items from each instrument correlate with each shared common factor based on factor loadings. While the former refers to items of the ICECAP-A and the respective HRQoL instrument that contribute to the same underlying latent factor, the latter describes the strength of this contribution. The correlation among common factors was also examined to explore the extent to which the instruments in each pairwise comparison measure separate but correlated factors. The robustness of results was examined by comparing the extent of overlap in the preferred factor model against the extent of overlap in alternative factor models for each pairwise comparison.

3 Results

Data from 6756 individuals were used in the analyses. Table 1 provides the characteristics of the study population for the combined sample and by country. Quota sampling was used in the MIC study and, therefore, the distributions of age, sex, and education level are similar across the countries. The presence of a chronic disease was self-reported by the majority (78%) of the study population.

3.1 'Preferred' Factor Models

Scree plots and the Kaiser criterion suggested a two-factor model for the EQ-5D-5L; a three-factor model for the 15D, HUI-3, and SF-6D; and a five-factor model for the AQoL-8D. In an attempt to improve model fit and interpretability, expansion of the number of factors was explored for all models. For the EQ-5D-5L and 15D, this resulted in an improvement in the model fit and factor structure with fewer cross-factor loadings, supporting the superiority of a three- and four-factor model, respectively. For the HUI-3, moving to a four-factor model improved model fit but resulted in a poorer factor structure and the three-factor model was retained as the preferred model. With regard to the SF-6D, a four-factor model was preferred because of a better model fit and a cleaner factor structure. A six-factor model was explored for the AQoL-8D but this did not improve interpretability of the factor structure and the five-factor model was retained. Results pertaining to the preferred factor model for each pairwise EFA are provided in Tables 2, 3, 4, 5 and 6.

Table 1 Characteristics of the study population [values are numbers (percentages) unless stated otherwise][a]

	Australia (N = 1341)	Canada (N = 1330)	Germany (N = 1269)	UK (N = 1356)	USA (N = 1460)	Total (N = 6756)
Age [mean (SD)]	53.91 (14.85)	48.74 (15.36)	49.18 (13.71)	51.55 (16.22)	52.00 (14.86)	51.11 (15.15)
Sex (female)	686 (48.8)	826 (62.1)	591 (46.6)	670 (49.4)	900 (61.6)	3707 (54.2)
Education						
High school	472 (35.2)	388 (29.2)	249 (19.6)	517 (38.1)	527 (36.1)	2193 (32.0)
Diploma or certificate or trade	469 (35.0)	633 (47.6)	698 (55.0)	409 (30.2)	428 (29.3)	2670 (39.0)
University	400 (29.8)	309 (23.2)	322 (25.2)	430 (31.7)	505 (34.6)	1982 (29.0)
Self-reported health condition						
Disease-free	265 (19.8)	328 (24.7)	260 (20.5)	298 (22.0)	321 (22.0)	1472 (21.5)
Asthma	141 (10.5)	138 (10.4)	147 (11.6)	150 (11.1)	150 (10.3)	726 (10.6)
Cancer	154 (11.5)	138 (10.4)	115 (9.1)	137 (10.1)	148 (10.1)	692 (10.1)
Depression	146 (10.9)	145 (10.9)	160 (12.6)	158 (11.7)	168 (11.5)	777 (11.4)
Diabetes mellitus	168 (12.5)	144 (10.8)	140 (11.0)	161 (11.9)	168 (11.5)	781 (11.4)
Hearing problems	155 (11.6)	144 (10.8)	136 (10.7)	126 (9.3)	156 (10.7)	717 (10.5)
Arthritis	163 (12.2)	139 (10.5)	159 (12.5)	159 (11.7)	179 (12.3)	799 (11.7)
Heart	149 (11.1)	154 (11.6)	152 (12.0)	167 (12.3)	170 (11.6)	792 (11.6)

ICECAP-A ICEpop CAPability measure for Adults, *SD* standard deviation

[a] Data from Norway are not used in this analysis because the ICECAP-A was not administered

3.2 Overlap with the ICECAP-A

Results suggest some degree of overlap between the ICECAP-A and the HRQoL instruments, although the extent of overlap varied across instruments. For the 15D EFA, two common factors were shared (Factors 2 and 4) [see Table 2]. In each case, ICECAP-A dimensions did not load strongly onto the respective shared factor [autonomy (0.337) on Factor 2 and stability (0.307) on Factor 4] and the shared factor mostly explained variance in the 15D items. All five ICECAP-A dimensions loaded strongly onto Factor 1, a factor that was not shared by any 15D items. However, Factor 1 was strongly correlated ($r = 0.714$) with Factor 4, which included the 15D items depression (0.841), distress (0.870), vitality (0.491), mental function (0.309), and sleeping (0.439).

The degree of overlap was much larger when comparing the ICECAP-A with the AQoL-8D. Three common factors (Factors 1–3) were shared by ICECAP-A and AQoL-8D items (see Table 3). Four ICECAP-A dimensions [stability (0.782), autonomy (0.345), achievement (0.634), and enjoyment (0.553)] and 18 AQoL-8D items loaded onto Factor 1. Factor 2 was shared by ICECAP-A autonomy (0.415) and 14 AQoL-8D items. Factor 3 included ICE-CAP-A attachment (0.682) and enjoyment (0.338), and six items of the AQoL-8D [social exclusion (0.307), close relationships (0.782), enjoy close relationships (0.842), pleasure (0.365), social isolation (0.338), and intimacy (0.581)]. Strong correlations were observed between

Factors 1 and 3 ($r = 0.643$), and Factors 1 and 4 ($r = 0.641$). Despite the strong correlation with Factor 1, Factor 4 was not a shared factor. Factor 4 comprised AQoL-8D items only, with the largest factor loadings being social exclusion (0.679) and social isolation (0.653).

The EQ-5D-5L shared two common factors with the ICECAP-A (Factors 1 and 3) [see Table 4]. Four ICECAP-A dimensions [stability (0.803), attachment (0.798), achievement (0.658) and enjoyment (0.826)] and EQ-5D-5L anxiety/depression (0.703) loaded onto Factor 1. Factor 3 was primarily represented by the ICECAP-A autonomy (0.657) and achievement (0.426), as well as EQ-5D-5L self-care (0.301). Whereas a moderate correlation was found between Factor 1 and Factor 3 ($r = 0.323$), a strong correlation ($r = 0.685$) was observed between Factor 3 and Factor 2, where Factor 2 comprised EQ-5D-5L items only.

The HUI-3 (Table 5) and SF-6D (Table 6) also shared two common factors with the ICECAP-A in the respective pairwise comparisons. All five ICECAP-A dimensions loaded onto the same factor as a single SF-6D item [energy (0.391)], and two HUI-3 items [emotion (0.895) and cognition (0.455)]. In both models, ICE-CAP-A autonomy cross-loaded onto a second factor that was shared by ambulation (0.883), dexterity (0.576), and pain (0.719) in the HUI-3 EFA, and five items from the physical functioning and role limitation dimensions in the SF-6D EFA. Moderate correlations were observed between the shared factors for the respective pairwise comparisons.

Table 2 EFA comparing the ICECAP-A with the 15D (four-factor model)[a]

		Rotated item loadings			
		Factor 1	Factor 2	Factor 3	Factor 4
ICECAP-A					
Stability		**<u>0.601</u>**	0.029	−0.021	**0.307**
Attachment		**<u>0.742</u>**	−0.165	0.061	0.090
Autonomy		**<u>0.497</u>**	**0.337**	0.064	−0.078
Achievement		**<u>0.762</u>**	0.224	−0.034	0.006
Enjoyment		**<u>0.802</u>**	0.019	0.010	0.109
15D					
Mobility		0.060	**<u>0.902</u>**	0.039	−0.167
Vision		0.042	0.147	**<u>0.468</u>**	−0.006
Hearing		−0.028	0.012	**<u>0.654</u>**	−0.121
Breathing		−0.030	**<u>0.563</u>**	0.064	0.145
Sleeping		−0.010	**0.374**	0.010	**<u>0.438</u>**
Eating		−0.003	0.280	**<u>0.648</u>**	−0.024
Speech		0.024	−0.091	**<u>0.815</u>**	0.094
Elimination		−0.112	**<u>0.371</u>**	0.277	0.172
Usual activities		0.152	**<u>0.812</u>**	−0.001	0.029
Mental function		0.085	0.085	**<u>0.474</u>**	**0.309**
Discomfort and symptoms		−0.088	**<u>0.763</u>**	−0.034	0.197
Depression		0.126	0.014	0.009	**<u>0.841</u>**
Distress		0.028	−0.016	0.056	**<u>0.870</u>**
Vitality		0.086	**0.472**	−0.002	**<u>0.491</u>**
Sexual activity		0.096	**<u>0.467</u>**	0.057	0.240
		Correlations among factors			
	Factor 1	1.000			
	Factor 2	0.463*	1.000		
	Factor 3	0.422*	0.562*	1.000	
	Factor 4	0.714*	0.428*	0.396*	1.000
RMSEA	0.042 [90% CI 0.040–0.044]				
CFI	0.991				
TLI	0.985				

CFI Comparative Fit Index, *CI* confidence interval, *EFA* exploratory factor analysis, *ICECAP-A* ICEpop CAPability measure for Adults, *RMSEA* root mean square error of approximation, *TLI* Tucker–Lewis Index

* Significant at 5% level

[a] Loadings >0.30 are presented in bold face. The highest loading for each item is underlined

3.3 Robustness of the Preferred Factor Models

Comparing the extent of overlap in the preferred factor models against alternative (larger or smaller) factor models identified differences in overlap for the pairwise analyses comprising the 15D, EQ-5D-5L, and HUI-3 (see Online Supplementary Material II–IV, respectively). For the 15D, a three-factor model suggested a higher degree of overlap with the ICECAP-A than the preferred four-factor model. For the three-factor 15D model, four 15D items [depression (0.691), distress (0.619), vitality (0.439), and sleeping (0.314)] and all five ICECAP-A dimensions were explained by Factor 1. For the EQ-5D-5L, a two-factor model confirmed the strong loading of anxiety/depression

onto Factor 1, but the remaining four EQ-5D-5L dimensions now shared a common factor with ICECAP-A autonomy, which loaded onto both common factors. Differences with regard to autonomy were also observed for the HUI-3. Unlike the preferred three-factor model, a four-factor model showed that autonomy loaded strongly on a factor that was not shared by any HUI-3 items.

4 Discussion

The ICECAP-A was developed to overcome perceived limitations associated with existing preference-based instruments that focus primarily (but not only) on health-

Table 3 EFA comparing the ICECAP-A with the AQoL-8D (five-factor model)[a]

	Rotated item loadings				
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
ICECAP-A					
Stability	**_0.782_**	0.033	0.084	−0.045	−0.156
Attachment	0.161	−0.045	**_0.682_**	0.107	−0.070
Autonomy	**0.345**	**_0.415_**	−0.098	0.080	−0.231
Achievement	**_0.634_**	0.245	0.081	−0.080	−0.264
Enjoyment	**_0.553_**	0.117	**0.338**	−0.044	−0.140
AQoL-8D					
Energy level	**_0.544_**	**0.390**	0.020	−0.049	0.034
Social exclusion	0.060	0.043	**0.307**	**_0.679_**	−0.004
Getting around	0.073	**_0.781_**	0.033	0.107	−0.037
Community role	−0.035	**_0.701_**	0.046	**0.348**	0.005
Sadness	**_0.742_**	−0.041	0.012	0.234	0.176
Frequency of pain	0.017	**_0.883_**	−0.001	−0.020	**0.523**
Confidence	**_0.721_**	−0.054	0.021	0.132	−0.132
Calm	**_0.765_**	0.010	−0.044	0.026	0.130
Family role	0.042	**_0.587_**	0.110	**0.346**	0.013
Close relationships	−0.002	0.002	**_0.782_**	0.228	0.054
Communication	0.002	0.201	0.218	**_0.301_**	−0.097
Sleep	**0.392**	**0.348**	0.007	0.015	0.196
Feeling worthless	**_0.653_**	−0.003	0.055	0.275	0.024
Anger	**_0.600_**	−0.074	0.038	0.072	0.185
Mobility	−0.079	**_0.937_**	−0.005	−0.005	0.004
Self-harm	**0.488**	−0.004	0.071	0.268	0.039
Enthusiasm	**_0.700_**	0.068	0.177	−0.056	−0.055
Worry	**_0.824_**	−0.036	−0.133	0.127	0.102
Self-care	0.018	**_0.796_**	0.018	0.124	−0.062
Happiness	**_0.735_**	−0.027	0.287	−0.032	0.039
Coping	**_0.775_**	0.072	−0.019	0.070	−0.121
Degree of pain	−0.009	**_0.920_**	−0.001	−0.065	**0.555**
Enjoy close relationships	0.018	−0.053	**_0.842_**	0.101	0.021
Pain interference	0.026	**_0.894_**	0.008	0.020	**0.395**
Pleasure	**_0.550_**	0.034	**0.365**	−0.035	0.011
Feeling burden	**0.367**	**0.333**	−0.085	**0.339**	−0.075
Contentment	**_0.723_**	0.043	0.235	−0.021	−0.045
Vision	0.004	**_0.389_**	0.170	−0.046	0.013
Control	**_0.809_**	0.013	0.019	0.045	−0.144
Household tasks	0.026	**_0.868_**	−0.068	0.056	−0.036
Social isolation	0.092	0.045	**0.338**	**_0.653_**	−0.031
Hearing	−0.133	**0.359**	0.210	0.014	−0.006
Depression	**_0.781_**	−0.019	0.011	0.189	0.174
Intimacy	0.199	0.085	**_0.581_**	0.034	0.043
Despair	**_0.733_**	0.019	−0.035	0.214	0.124
	Correlations among factors				
Factor 1	1.000				
Factor 2	0.494*	1.000			
Factor 3	0.643*	0.354*	1.000		
Factor 4	0.641*	0.301*	0.336*	1.000	
Factor 5	0.000	−0.222*	−0.081*	−0.048*	1.000

Table 3 continued

| | Rotated item loadings | | | | |
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
RMSEA	0.061 [90% CI 0.061−0.062]				
CFI	0.974				
TLI	0.965				

AQoL-8D Assessment of Quality of Life 8-dimension, *CFI* Comparative Fit Index, *CI* confidence interval, *EFA* exploratory factor analysis, *ICECAP-A* ICEpop CAPability measure for Adults, *RMSEA* root mean square error of approximation, *TLI* Tucker–Lewis Index

* Significant at 5% level

[a] Loadings >0.30 are presented in bold face. The highest loading for each item is underlined

Table 4 EFA comparing the ICECAP-A with the EQ-5D-5L (three-factor model)[a]

| | Rotated item loadings | | |
	Factor 1	Factor 2	Factor 3
ICECAP-A			
Stability	**0.803**	0.001	0.147
Attachment	**0.798**	−0.054	−0.003
Autonomy	0.279	0.020	**0.657**
Achievement	**0.658**	−0.015	**0.426**
Enjoyment	**0.826**	0.028	0.132
EQ-5D-5L			
Mobility	−0.137	**0.981**	0.003
Self-care	−0.008	**0.649**	**0.301**
Usual activities	0.041	**0.824**	0.112
Pain/discomfort	0.002	**0.981**	−0.234
Anxiety/depression	**0.703**	0.169	−0.020
	Correlation among factors		
Factor 1	1.000		
Factor 2	0.461*	1.000	
Factor 3	0.323*	0.685*	1.000
RMSEA	0.074 [90% CI 0.069–0.078]		
CFI	0.993		
TLI	0.983		

CFI Comparative Fit Index, *CI* confidence interval, *EFA* exploratory factor analysis, *ICECAP-A* ICEpop CAPability measure for Adults, *RMSEA* root mean square error of approximation, *TLI* Tucker–Lewis Index

* Significant at 5% level

[a] Loadings >0.30 are presented in bold face. The highest loading for each item is underlined

Table 5 EFA comparing the ICECAP-A with the HUI-3 (three-factor model)[a]

| | Rotated item loadings | | |
	Factor 1	Factor 2	Factor 3
ICECAP-A			
Stability	**0.860**	0.020	−0.096
Attachment	**0.849**	−0.217	0.045
Autonomy	**0.448**	**0.368**	−0.013
Achievement	**0.762**	0.233	−0.106
Enjoyment	**0.912**	−0.023	−0.016
HUI-3			
Vision	−0.016	0.233	**0.324**
Hearing	−0.014	0.064	**0.617**
Speech	0.299	−0.007	**0.704**
Ambulation	−0.023	**0.883**	0.090
Dexterity	0.058	**0.576**	0.249
Emotion	**0.895**	−0.078	0.027
Cognition	**0.455**	0.121	**0.343**
Pain	0.107	**0.719**	0.022
	Correlations among factors		
Factor 1	1.000		
Factor 2	0.460*	1.000	
Factor 3	0.223*	0.282*	1.000
RMSEA	0.054 [90% CI 0.051–0.058]		
CFI	0.990		
TLI	0.981		

CFI Comparative Fit Index, *CI* confidence interval, *EFA* exploratory factor analysis, *HUI-3* Health Utilities Index Mark 3, *ICECAP-A* ICEpop CAPability measure for Adults, *RMSEA* root mean square error of approximation, *TLI* Tucker–Lewis Index

* Significant at 5% level

[a] Loadings >0.30 are presented in bold face. The highest loading for each item is underlined

related aspects of quality of life. Our analyses have shown that the ICECAP-A provides information over and above that garnered from several commonly used preference-based HRQoL instruments. However, the level of overlap with the ICECAP-A varied across instruments. Compared with other preference-based HRQoL instruments, more common factors were identified between the ICECAP-A and AQoL-8D. Based on item loadings, these three

common factors can be described as reflecting aspects of wellbeing (Factor 1), physical health (Factor 2), and relationships (Factor 3). Some but not all of these common factors emerged from other pairwise comparisons. The third factor, relationships, was not identified when

Table 6 EFA comparing the ICECAP-A with the SF-6D (four-factor model)[a]

	Rotated item loadings			
	Factor 1	Factor 2	Factor 3	Factor 4
ICECAP-A				
Stability	**0.726**	−0.034	−0.150	0.017
Attachment	**0.780**	0.130	−0.004	0.066
Autonomy	**0.395**	**−0.423**	−0.102	−0.077
Achievement	**0.777**	−0.293	0.023	−0.068
Enjoyment	**0.859**	0.009	−0.010	0.094
SF-6D				
Vigorous activities	−0.088	**0.738**	−0.202	−0.171
Moderate activities	−0.014	**0.959**	−0.044	−0.030
Bathing or dressing oneself	0.036	**0.735**	0.188	−0.002
Limited in kind of work or other activities	0.098	**0.586**	0.288	−0.225
Accomplished less than you would like	−0.025	**0.311**	**0.656**	0.021
Frequency health problems interfered with social activities	−0.082	0.279	**0.512**	−0.187
Intensity of bodily pain	0.033	−0.075	0.030	**0.809**
Extent pain interfered with normal work	0.001	−0.015	−0.057	**0.968**
Been very nervous	−0.058	−0.027	**0.743**	−0.004
Felt downhearted and depressed	−0.230	−0.025	**0.729**	−0.015
Have a lot of energy	**0.391**	−0.250	−0.097	0.166
	Correlations among factors			
Factor 1	1.000			
Factor 2	−0.372*	1.000		
Factor 3	−0.627*	0.312*	1.000	
Factor 4	0.374*	−0.734*	−0.387*	1.000
RMSEA	0.075 [90% CI 0.073–0.078]			
CFI	0.985			
TLI	0.971			

CFI Comparative Fit Index, *EFA* exploratory factor analysis, *ICECAP-A* ICEpop CAPability measure for Adults, *RMSEA* root mean square error of approximation, *TLI* Tucker–Lewis Index

* Significant at 5% level

[a] Loadings >0.30 are presented in bold face. The highest loading for each item is underlined

comparing the ICECAP-A with the SF-6D, EQ-5D-5L, or HUI-3. Only one factor explained the overlap with the 15D, which was related to aspects of physical health.

Compared with other literature, similar results were identified by recent studies that conducted an EFA with the ICECAP-A and the EQ-5D-3L [28], as well as with the ICECAP-O and EQ-5D-3L [27]. In these studies, the respective ICECAP instrument and the EQ-5D-3L measured two separate but correlated factors, with the majority of the EQ-5D-3L items loading onto one factor and the majority of the respective ICECAP items loading onto the second. Only EQ-5D-3L anxiety/depression loaded strongly onto the same factor as four dimensions of the ICECAP-A (stability, attachment, achievement, and enjoyment) and ICECAP-O (attachment, security, role, and enjoyment), while ICECAP-A autonomy and ICECAP-O control loaded moderately onto both factors. The authors of

the two previous EFA studies conclude that the EQ-5D-3L and ICECAP instruments provide complementary information and, therefore, should not be treated as substitute outcome measures. Specific to the EQ-5D-5L, these findings are confirmed by the current study owing to the relatively minimal overlap observed with the ICECAP-A.

Similar conclusions can be drawn about the 15D, HUI-3 and SF-6D, where relatively few items loaded onto the same common factor(s) as the ICECAP-A items. In contrast, the AQoL-8D provided good coverage of the three factors it shared with the ICECAP-A, with 18 AQoL-8D items loading on Factor 1 (wellbeing factor), 14 AQoL-8D items loading on Factor 2 (physical health factor), and six AQoL-8D items loading on Factor 3 (relationships factor). As a note of caution, the overlap observed between the AQoL-8D and ICECAP-A does not endorse any suggestion that these two measures are substitutes.

The observed differences in overlap across instruments may be the result, inter alia, of differences in the framing of items (e.g., question formats, response options, recall time, etc.), based on evidence of previous comparative studies of preference-based HRQoL instruments [29, 31, 49]. The combination of different health issues within a single item [e.g., anxiety and depression (EQ-5D-5L); downhearted and depressed (SF-6D); and sad, melancholic, or depressed (15D)] may also contribute to the differences observed between instruments. More generally, the fact that the instruments included in this study differ in the way they conceptualize HRQoL [32], and in their coverage of domains to define health states, is likely to be a primary reason for the variation in study findings [29, 30]. To illustrate, compared with other instruments, the AQoL-8D has a strong focus on the psycho-social domain (25 out of 35 items) and contains questions in its descriptive system that have the greatest ability to capture the concept of capability wellbeing, or wellbeing in general. As has been shown in a previous publication using data from the MIC study, which compared three subjective wellbeing instruments (Satisfaction with Life Scale, Personal Wellbeing Index, and the Integrated Household Survey of the Office for National Statistics) with preference-based HRQoL instruments, the AQoL-8D accounted for variation in subjective wellbeing to a greater extent than the other preference-based HRQoL instruments [50].

4.1 Implications and Directions for Further Research

This study has shown that the ICECAP-A, when compared directly with the 15D, EQ-5D-5L, HUI-3, and SF-6D, provides additional complementary information in terms of the impact of an intervention on an individual's capability wellbeing. Recent studies have demonstrated that the choice of outcome measure for economic evaluation, i.e., selecting a capability measure *or* a HRQoL measure, is not a trivial issue [51, 52]. In an economic evaluation of an integrated care model for frail seniors, Makai and colleagues found the intervention had a higher probability of being cost effective when using the ICECAP-O when compared with use of the EQ-5D-3L [51]. This direct comparison of cost-effectiveness findings was made possible because (1) ICECAP-O responses were used to define 'capability QALYs' and (2) the same range of willingness to pay (WTP) values was applied in the analysis of capability QALYs and QALYs derived from EQ-5D-3L responses. Despite the use of identical economic evaluation approaches, Makai and colleagues go on to highlight that there are no estimates of WTP for a capability QALY, and state that it is unlikely that valid comparisons can be made between the ICECAP-O and EQ-5D-3L at a given level of

WTP. A second example examined the cost effectiveness of psychological interventions for drug addiction [52], concluding that under the health maximization principle (using EQ-5D-5L), the results yielded different treatment recommendations when compared with the application of the 'sufficient capability' approach developed by Mitchell and colleagues (using ICECAP-A) [53].

Although methodologies to operationalize the use of ICECAP instruments in economic evaluation are still in their infancy [53], findings such as those in the above examples support the use of ICECAP instruments alongside preference-based HRQoL instruments to triangulate results and evaluate the robustness of conclusions regarding cost effectiveness. However, the use of different metrics to value different healthcare interventions raises questions about the objective for resource allocation decisions in healthcare [54], e.g., does health or wellbeing (or both) enter the objective function, and is the form of this function consistent with the current emphasis on maximization (rather than sufficiency)? To answer this question, further research is needed to determine whether a society is willing to sacrifice health outcomes for improvements in dimensions of wellbeing.

The use of ICECAP instruments within the current QALY-based paradigm for economic evaluation also requires further attention in health economics research. As mentioned above, the ICECAP-A is anchored on a 'full capability' and 'no capability' scale and the instrument was not intended to be used within the QALY framework. Recent advances in this area have proposed to adjust the ICECAP-A for time to enable the assessment of gains in terms of 'years of full capability equivalence' [24], and an approach that focuses on the objective of achieving 'sufficient capability' [53]. Outside the ICECAP instruments, Cookson suggested an application of the capability approach to economic evaluation by re-interpreting the QALY, referred to as the 'capability QALY' [55]. Cookson argued that, in practice, HRQoL instruments incorporate some elements of capability because health affects an individual's freedom to choose non-health activities. Compared with the 'health QALY', this operationalization of the 'capability QALY' represents individuals' entire wellbeing (not just the health component), and, therefore, reflects the value of the capability set. Concerns over using preference-based HRQoL instruments as the base of a capability QALY because they may neglect non-health dimensions of wellbeing led Cookson to conclude that, "… the QALY approach is compatible with the capability approach only insofar as the health state descriptive systems used for generating QALYs pay close attention to proxy capability variables that cover a wide range of health and non-health dimensions of wellbeing" [56]. Results from the current study suggest the AQoL-8D could be a measure that best fits

Cookson's notion of a capability QALY because of the overlap with the ICECAP-A and the presence of non-health items in the AQoL-8D descriptive system.

In a recent review, Karimi and colleagues conclude that existing capability measures (including ICECAP instruments) have important limitations because they do not elicit capability as originally intended in the capability approach [57]. Accordingly, if the added value of capability instruments in health economics is based solely on broadening the evaluative space to extend beyond a narrow focus on health, our findings provide evidence that such benefits can be potentially captured, to some degree, by the AQoL-8D (i.e., not only through the aggregation of outcomes collected by 'complementary' health-related and capability measures). However, as alluded to earlier, our findings do not imply the AQoL-8D and ICECAP-A are interchangeable instruments. Further work is needed to build on these findings and explore unanswered questions, such as whether individuals are able to distinguish between their capabilities and functionings, and the comparative performance of the ICECAP-A and AQoL-8D with regard to capturing the wellbeing impacts of interventions in different clinical contexts. It is also important to note that ICECAP instruments are not the only capability measures that could be combined with QALYs derived from HRQoL instruments to provide a broader assessment of the benefit of interventions. For example, the Adult Social Care Outcomes Toolkit (ASCOT) [10] is designed to capture information about an individual's social care-related quality of life and further research is needed to explore the relationships (including overlap) between the ASCOT, ICECAP instruments, and preference-based HRQoL instruments.

4.2 Strengths and Limitations

A major strength of this study is the inclusion of multiple preference-based instruments. While previous studies explored the overlap between ICECAP instruments and the EQ-5D-3L (using much smaller samples [27, 28]), this study provides EFA results comparing the ICECAP-A with five preference-based HRQoL instruments. Conducting an EFA that uses data from the descriptive systems only (i.e., item-level response data) is a further strength because there is no reliance on country-specific index scores, where variations across national valuation studies could influence the results [58, 59]. Given that 'overlap' between instruments can be explored within the descriptive systems or health state valuations, this item-level analysis complements previous work that used correlation analyses and regression-based techniques to assess index scores from the MIC study [60]. The analysis also addressed the potential problem of factor under- or over-extraction by investigating alternative factor models to examine the robustness of the

'preferred' factor models [61]. Potential limitations associated with using data from a multinational survey include issues with the validity of instrument translations and the representation of the respective populations (for example, participants were required to have Internet access). Survey bias resulting from the repetition of similar items should also be acknowledged owing to the administration of seven preference-based HRQoL instruments.

5 Conclusion

The ICECAP-A has the potential to capture benefits of interventions and treatments that go beyond those measured by many of the traditional health-focused preference-based instruments, such as the 15D, EQ-5D-5L, HUI-3, and SF-6D. Substantial overlap was observed between the ICECAP-A and AQoL-8D. Researchers and decision makers should be aware that there is a risk of double counting when using the ICECAP-A as a complementary measure, but the level of such a risk varies depending on the choice of HRQoL measure. Further investigations are needed to explore the extent and implications of double counting, particularly when applying the ICECAP-A alongside the AQoL-8D.

References

1. Drummond M, Sculpher M, Claxton K, et al. Methods for the economic evaluation of health care programmes. 4th ed. Oxford: Oxford University Press; 2015.

2. Weinstein MC, Torrance G, McGuire A. QALYs: the basics. Value Health. 2009;12(Suppl. 1):S5–9.

3. CADTH. Guidelines for the economic evaluation of health technologies: Canada, 3rd edn. Ottawa: CADTH; 2006.

4. NICE. Guide to the methods of technology appraisal. London: NICE; 2013.

5. Nederland Zorginstituut. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. Diemen: Zorginstituut Nederland; 2016.

6. Mooney G. QALYs: are they enough? A health economist's perspective. J Med Ethics. 1989;15(3):148–52.

7. Mooney G. Beyond health outcomes: the benefits of health care. Health Care Anal. 1998;6(2):99–105.

8. Ryan M, Shackley P. Assessing the benefits of health care: how far should we go? Qual Health Care. 1995;4(3):207–13.

9. Lorgelly PK, Lawson KD, Fenwick EA, Briggs AH. Outcome measurement in economic evaluations of public health interventions: a role for the capability approach? Int J Environ Res Public Health. 2010;7(5):2274–89.

10. Netten A, Burge P, Malley J, et al. Outcomes of social care for adults: developing a preference-weighted measure. Health Technol Assess. 2012;16(16):1–166.

11. Connell J, O'Cathain A, Brazier J. Measuring quality of life in mental health: are we asking the right questions? Soc Sci Med. 2014;120:12–20.

12. Coast J. Strategies for the economic evaluation of end-of-life care: making a case for the capability approach. Expert Rev Pharmacoecon Outcomes Res. 2014;14(4):473–82.

13. Makai P, Brouwer WBF, Koopmanschap MA, et al. Quality of life instruments for economic evaluations in health and social care for older people: a systematic review. Soc Sci Med. 2014;102:83–93.

14. Dodge R, Daly AP, Huyton J, Sanders LD. The challenge of defining wellbeing. Int J Wellbeing. 2012;2(3):222–35.

15. Ryff D, Singer B. Know thyself and become what you are: a eudaimonic approach to psychological well-being. J Happiness Stud. 2008;9(1):13–39.

16. Diener E, Suh EM, Lucas RE, Smith HL. Subjective well-being: three decades of progress. Psychol Bull. 1999;125(2):276–302.

17. Sen A. Capability and well-being, in the quality of life. Nussbaum M, Sen A, editors. Oxford: Oxford University Press; 1993.

18. Bleichrodt H, Quiggin J. Capabilities as menus: a non-welfarist basis for QALY evaluation. J Health Econ. 2013;32(1):128–37.

19. Al-Janabi H, Flynn T, Coast J. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. Qual Life Res. 2012;21(1):167–76.

20. Lorgelly PK. Choice of outcome measure in an economic evaluation: a potential role for the capability approach. Pharmacoeconomics. 2015;33(8):849–55.

21. Coast J, Peters TJ, Natarajan L, et al. An assessment of the construct validity of the descriptive system for the ICECAP capability measure for older people. Qual Life Res. 2008;17(7):967–76.

22. Sutton EJ, Coast J. Development of a supportive care measure for economic evaluation of end-of-life care using qualitative methods. Palliat Med. 2014;28(2):151–7.

23. NICE. Developing NICE guidelines: the manual. London: NICE; 2014.

24. Flynn TN, Huynh E, Peters TJ, et al. Scoring the ICECAP-A capability instrument: estimation of a UK general population tariff. Health Econ. 2015;24(3):258–69.

25. Brooks R. EuroQol: the current state of play. Health Policy. 1996;37(1):53–72.

26. Herdman M, Gudex C, Lloyd A, 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.

27. 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.

28. Keeley T, Coast J, Nicholls E, et al. 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):36.

29. Richardson J, Iezzi A, Khan MA. Why do multi-attribute utility instruments produce different utilities: the relative importance of the descriptive systems, scale and 'micro-utility' effects. Qual Life Res. 2015;24(8):2045–53.

30. Richardson J, McKie J, Bariola E. Multiattribute utility instruments and their use. In: Culyer AJ, editor. Encyclopedia of health economics. San Diego: Elsevier; 2014. p. 341–57.

31. Whitehurst DG, Norman R, Brazier JE, Viney R. Comparison of contemporaneous EQ-5D and SF-6D responses using scoring algorithms derived from similar valuation exercises. Value Health. 2014;17(5):570–7.

32. Karimi M, Brazier J. Health, health-related quality of life, and quality of life: what is the difference? Pharmacoeconomics. 2016;34(7):645–9.

33. Richardson J, Iezzi A, Maxwell A. Cross-national comparison of twelve quality of life instruments. MIC paper 1: background, questions, instruments. Research paper 76. Melbourne (VIC): Centre for Health Economics, Monash University; 2012. Available from: http://www.aqol.com.au/papers/researchpaper76.pdf. Accessed 28 Jan 2017.

34. Sintonen H. The 15D instrument of health-related quality of life: properties and applications. Ann Med. 2001;33(5):328–36.

35. Hawthorne G, Richardson J, Osborne R. The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of health-related quality of life. Qual Life Res. 1999;8(3):209–24.

36. Richardson J, Iezzi A, Khan MA, et al. Data used in the development of the AQoL-8D (PsyQoL) quality of life instrument. Melbourne: Centre for Health Economics, Monash University; 2009.

37. Feeny D, Furlong W, Torrance GW, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. Med Care. 2002;40(2):113–28.

38. Seiber WJ, Groessl EJ, David KM, et al. Quality of Well Being Self Administered (QWB-SA) Scale: user's manual. 2008. Available from: https://hoap.ucsd.edu/qwb-info/QWB-Manual.pdf. Accessed 28 Jan 2017.

39. Ware JE Jr, Kosinski M, Bjorner JB, et al. SF-36v2® Health Survey: administration guide for clinical trial investigators. Lincoln: Quality Metric Incorporated; 2008.

40. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ. 2002;21(2):271–92.

41. Perreault WD. Controlling order-effect bias. Public Opin Q. 1976;39(4):544–51.

42. Muthén LK, Muthén BO. Mplus user's guide, 7th ed (1998–2015). Los Angeles: Muthén & Muthén; 2015.

43. Fabrigar LR, Wegener DT. Understanding statistics: exploratory factor analysis. New York: Oxford University Press; 2012.

44. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: Erlbaum; 1988.

45. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. Psychol Methods. 1999;4(3):272–99.

46. Geiser C. Data analysis with Mplus. New York: Guilford Press; 2013.

47. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling. 1999;6(1):1–55.

48. Costello AB, Osborne JW. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Pract Assess Res Eval. 2005;10(7):1–9.

49. Whitehurst DG, Bryan S. Another study showing that two preference-based measures of health-related quality of life (EQ-5D and SF-6D) are not interchangeable. But why should we expect them to be? Value Health. 2011;14(4):531–8.

50. Richardson J, Chen G, Khan MA, Iezzi A. Can multi-attribute utility instruments adequately account for subjective well-being? Med Decis Mak. 2015;35(3):292–304.

51. Makai P, Looman W, Adang E, et al. Cost-effectiveness of integrated care in frail elderly using the ICECAP-O and EQ-5D: does choice of instrument matter? Eur J Health Econ. 2015;16(4):437–50.

52. Goranitis I, Coast J, Day E, et al. Maximizing health or sufficient capability in economic evaluation? A methodological experiment of treatment for drug addiction. Med Decis Mak. 2016. doi:10.1177/0272989X16678844.

53. Mitchell PM, Roberts TE, Barton PM, Coast J. Assessing sufficient capability: a new approach to economic evaluation. Soc Sci Med. 2015;139:71–9.

54. Mitchell PM, Al-Janabi H, Richardson J, et al. The relative impacts of disease on health status and capability wellbeing: a multi-country study. PLoS One. 2015;10(12):e0143590.

55. Cookson R. QALYs, and the capability approach. Health Econ. 2005;14(8):817–29.

56. Cookson R. QALYs and capabilities: a response to Anand. Health Econ. 2005;14(12):1287–9.

57. Karimi M, Brazier J, Basarir H. The capability approach: a critical review of its application in health economics. Value Health. 2016;19(6):795–9.

58. Engel L, Bansback N, Bryan S, et al. Exclusion criteria in national health state valuation studies: a systematic review. Med Decis Mak. 2016;36(7):798–810.

59. Xie F, Gaebel K, Perampaladas K, et al. Comparing EQ-5D valuation studies: a systematic review and methodological reporting checklist. Med Decis Mak. 2013;34(1):8–20.

60. Mitchell PM, Venkatapuram S, Richardson J, Iezzi A, Coast J. Are quality-adjusted life years a good proxy measure of individual capabilities? Pharmacoeconomics. 2017. doi:10.1007/s40273-017-0495-3.

61. Prieto L, Alonso J, Lamarca R. Classical test theory versus rasch analysis for quality of life questionnaire reduction. Health Qual Life Outcomes. 2003;1:27.