Quantifying the Relationship between Capability and Health in Older People: Can't Map, Won't Map

Matthew Franklin, PhD, Katherine Payne, PhD, Rachel A. Elliott, PhD

Background. *Intuitively, health and capability are distinct but linked concepts. This study aimed to quantify the link between a measure of health status (EQ-5D-3L) and capability (ICECAP-O) using regression-based methods.* **Methods.** *EQ-5D-3L and ICECAP-O data were collected from a sample of older people (*n = 584*), aged over 65 years, requiring a hospital visit and/or care home resident, and recruited to one of 3 studies forming the Medical Crisis in Older People (MCOP) program in England. The link of EQ-5D-3L with 1) ICECAP-O tariff scores were estimated using ordinary least squares (OLS) or censored least absolute deviation (CLAD) regression models; and 2) ICECAP-O domain scores was estimated using multinomial logistic (MNL) regression. Mean absolute error (MAE), root mean squared error (RMSE), absolute difference (AD) between mean observed and estimated values, and the R^2 statistic were used to judge model performance.* **Results.** *In this sample of older people (*n = 584*), higher scores on the EQ-5D-3L were shown to be linked with higher ICECAP-O scores when using linear regression. An OLS-regression model was identified to be the best performing model with the lowest error statistics (AD = 0.0000; MAE = 0.1208; MSE = 0.1626) and highest goodness of fit (R^2 = 0.3532); model performance was poor when predicting the lower ICECAP-O tariff scores. The three domains of the EQ-5D-3L showing a statistically significant quantifiable link with the ICECAP-O tariff score were self-care, usual activities, and anxiety/depression.* **Conclusion.** *A quantifiable, but weak, link between health (EQ-5D-3L) and capability (ICECAP-O) was identified. The findings from this study add further support that the ICECAP-O is providing complimentary information to the EQ-5D-3L. Mapping between the 2 measures is not advisable and the measures should not be used as direct substitutes to capture the impact of interventions in economic evaluations.* **Key words:** *regression analysis; EQ-5D-3L; ICECAP-O; capability; health status; quality of life.* **(Med Decis Making XXXX;XX:xx–xx)**

Health status, captured using the quality-adjusted life year (QALY) measure,[1,2] and quantified using generic preference-based measures,[3] has become the commonly recommended outcome to value the consequences of healthcare programs in economic evaluations.[4] Guidance on how to perform economic evaluations to inform resource allocation within healthcare systems, such as the National Institute for Health and Care Excellence (NICE) methods guide for technology appraisal,[5] has led to a focus on using the EQ-5D[5,6] as the relevant metric to quantify the health consequences of healthcare programmes.[7–9]

Relevant consequences broader than health status, as defined by the EQ-5D,[3] have been suggested.[10–13]

Moving beyond health status may be particularly relevant when valuing the consequences of complex[12] and social care interventions.[14] An alternative consequence is the notion of capability introduced as a concept by Sen.[15–17] Sen suggested that the relevant objective for policy-makers should be based on a person's ability to "do" or "be."[16–20] Lorgelly[21] and Coast et al.[22] provide an overview of measures that aim to quantify capability and their potential use in economic evaluation.

The suite of ICECAP measures are considered useful, as they have associated preference weights, making them viable measures of capability for use in economic evaluations. The ICECAP measures were the first capability-based measures designed for use in economic evaluations.[23,24] The ICECAP suite includes the ICECAP-O (the full measure can be downloaded from the ICECAP [University of Birmingham] website[25]) for use in older people[26] and ICECAP-A for adults.[27] NICE guidance

currently recommends the ICECAP measures as an option in economic evaluation of social care interventions.[28] The developers suggest that the ICECAP measures capture the impact of a broader aspect of quality of life rather than only health status as captured by the EQ-5D. Differences are apparent in both the conceptualization and practical design of the ICECAP and EQ-5D measures, respectively. Conceptual differences are apparent in the question wording to bring the focus of ICECAP in line with Sen's theoretical underpinning using the terminology of "ability" rather than the focus on "functioning" seen in EQ-5D. Two practical design differences are apparent: the severity levels used (4 levels for ICECAP v. 3 for EQ-5D-3L) and different tariff scoring scales. Both measures use the same upper bound score of 1, representing either "full capability" or "perfect health." The measures differ in how the lower bound is anchored at zero. For ICECAP, zero represents "no capability." For EQ-5D, zero represents a state equivalent to dead but it is also possible to have states worse than dead with negative scores. Thus, the 2 measures operate on different scales, with implications for their direct comparison. There is currently no definitive guidance on how to use ICECAP measures in economic evaluation (i.e., should they be used in a QALY-based approach or otherwise?). Even though the ICECAP measures are preference-based, this does not mean they could be used to quantify a QALY; although, at least one study has chosen this path.[29] Relatedly, it

is not clear that the QALY is an appropriate endpoint for evaluating capability.[30] This has led to work on an alternative capability-based method for economic evaluation.[31] Therefore, research into how capability should be operationalized as part of an economic evaluation should be considered as "ongoing."

Sen was purposely vague in his definition of capability.[17,32] Nussbaum suggested a need for more specificity in what is meant by capability.[33,34] Grewal and colleagues[35] suggested that "It is not poor health in itself, which reduces quality of life, but the influence of that poor health upon each informant's ability to, say, be independent, that is important". In this context, health is perceived as a conversion factor for capability. Capability is viewed as the objective end-goal for patients when receiving healthcare rather than only health status. Available measures of health status (EQ-5D-3L) and capability (ICECAP-O) have been argued to be conceptually different,[21] and a previous study has suggested that the 2 measures offer complementary information (rather than the ICECAP-O acting as a substitute to the EQ-5D-3L, because it captures essentially the same information).[36] However, it seems logical, based on their conceptual underpinning, that health and capability could be linked in some quantifiable way and, therefore, a change in health captured by the EQ-5D-3L may be associated with a change in capability captured by the ICECAP-O.[37] One study has identified that capability (ICECAP-O) was strongly and positively associated with health status (EQ-5D-3L); however, that study focused on a specific population of older people receiving post-acute rehabilitation care and was a relatively small sample ($n = 82$).[38] The current study aimed to build on this work and quantify the link between the EQ-5D-3L and the ICECAP-O to quantify the relationship (association) between the constructs of the measures, health, and capability, respectively.

METHODS

This study used regression-based statistical methods to quantify the link between the EQ-5D-3L and ICECAP-O, informed by published guidance and recommended reporting standards for mapping studies (also called "cross-walking" or "transfer to utility").[39–44] The ICECAP-O was used as the target measure (response variable) for this assessment so that the descriptive results suggest to what extent a

Address correspondence to Matthew Franklin, Health Economics and Decision Science (HEDS), School of Health and Related Research (ScHARR), University of Sheffield, West Court, 1 Mappin Street, Sheffield S1 4DT, UK; Telephone: +44 (0)114 222 4226; e-mail: matt.franklin@sheffield.ac.uk.

change in health is associated with a change in capability. This approach is consistent with the conceptual idea that health is a conversion factor for capability rather than vice versa (i.e., using the EQ-5D-3L as the target measure).

Study Sample

Data were collected from 3 observational cohort studies that formed the Medical Crises in Older People (MCOP) program: Acute Medical Outcomes Study (AMOS); Better Mental Health (BMH) study; and Care Home Outcomes Study (CHOS).

AMOS[45,46] included older people (70 y and over) admitted to hospital and discharged within 72 h from an Acute Medical Unit (AMU), in Nottingham or Leicester, England. Baseline data ($n = 667$) were collected at recruitment and follow-up data collected 90 d post-recruitment. Patients were excluded if staff advised against approaching the patient, or if neither the patient nor the carer could communicate in English sufficiently to complete baseline assessments. Patients who lacked the mental capacity to consent to study participation were recruited provided a responsible physician gave permission.

BMH[47,48] included older people (70 y and over) with a co-morbid mental health problem; with an unplanned admission to an acute general hospital in Nottingham, England, lasting 2 or more days to one of 12 named wards (trauma orthopedic, acute geriatric medical, general medical); and who had been screened using brief tests of cognition,[49] depression,[50] anxiety,[51] or alcohol misuse.[52] Baseline data ($n = 250$) were collected at recruitment and follow-up data collected 180 d post-recruitment. Patients viewed to have sufficient mental capacity gave written informed consent. Patients viewed to lack sufficient mental capacity were recruited provided a family member or carer gave permission.

CHOS[53] included older people (65 y and over) living in either a residential or nursing care home. Eleven (6 residential and 5 nursing) care homes within the Nottinghamshire catchment area were recruited. Baseline data ($n = 227$) were collected at recruitment and follow-up data collected at 180 d post-recruitment. Outcomes (EQ-5D-3L; ICECAP-O) were recorded using proxy and self-reported approaches but the study did not record when either approach was used. Care home managers determined which residents had the mental capacity to consent to participate, defined against the criteria in the English Mental Capacity Act.[54] If residents lacked capacity, a consultee was identified and, if they were in favor of proceeding, the resident was enrolled.

Dataset

Data collected in all 3 studies included: age, gender, whether living in a care home (nursing or residential), EQ-5D-3L, and ICECAP-O (follow-up only). Appendix S1 details the outcome measures collected. The analysis used ICECAP-O and EQ-5D-3L data collected between April 2009 and February 2011.

The ICECAP-O comprises 5 attributes reflecting capability (attachment, security, role, enjoyment, and control), each of which has 4 levels (the full measure can be downloaded from the ICECAP [University of Birmingham] website[25]). The ICECAP-O tariff score is anchored between 0 (no capability) and 1 (full capability). The preference-based scoring tariff for the ICECAP-O[26] was quantified using the best–worst scaling technique.[55] The measure, which was specifically designed for older people, has proven construct,[37, 56-58] convergent and discriminant,[59-61] and face[62] validity.

The EQ-5D-3L comprises 5 attributes reflecting health status (mobility, self-care, usual activities, pain/discomfort, anxiety/depression), each of which has 3 levels (the full measure can be downloaded from the EuroQoL website[63]). The UK preference-based tariff score of the EQ-5D-3L ranges from -0.594 (a state worse than dead) to 1 (perfect health). The value of zero is representative of the state of dead. The UK's preference-based scoring tariff for the EQ-5D-3L[64] was quantified using the time–trade-off technique.[65] A structured review of the generic self-assessed health instruments suggested that there is good evidence of validity (construct, convergent and discriminant) for the use of the EQ-5D-3L in older people.[66]

Estimation Dataset

The estimation dataset (hereafter called "MCOP") used data from all 3 studies combined. The appropriateness of using this combined dataset was tested using a linear regression model to assess the degree of association among the data coming from 1 of the 3 study samples and the ICECAP-O tariff score (see Appendix S2). There was concern that cognitive ability may need to be accounted for within this analysis, as it could be associated with a

given response to the ICECAP-O. A linear regression model was used to test the degree of association between cognitive ability (defined by Mini Mental State Examination [MMSE] score as a continuous or discrete dummy variable[67,68]) and the ICECAP-O tariff score. This analysis indicated that being in a particular study sample or cognitive ability (MMSE score) did not have a statistically significant association with the ICECAP-O tariff score.

Older people are on a continuum often characterized by aspects such as co-morbidities, physical and mental health, and cognitive ability. However, the 3 study samples are artificial groupings defined by place of recruitment and eligibility criteria but, overall, probably represent the continuum of older people more than any one study sample on its own (see also Graham et al.[69] and Carlo et al.,[70] which describe this continuum in the case of prevalence of cognitive ability in older people with and without dementia). Therefore, the results of the statistical analysis to inform combining these groups were not unexpected.

Data Analysis

All analyses were carried out using Stata version 11.[71] The estimation dataset (MCOP) and data from the 3 studies were analyzed using descriptive statistics and the distributions of the EQ-5D-3L and ICECAP-O data. Regression techniques were used to analyze the estimation dataset (MCOP) to quantify the strength of association between the EQ-5D-3L and ICECAP-O.

To inform the choice of regression models, the distribution of the ICECAP-O and EQ-5D-3L were assessed. Figure 1 shows the distribution of the EQ-5D-3L and ICECAP-O within the estimation dataset and illustrates that both measures had skewed and bimodal or multimodal tariff score distributions. Importantly, both measures were identified to have non-normally distributed tariff scores, and ceiling effects were apparent.

Based on this result, 3 types of regression models were investigated to identify which was most appropriate in terms of taking into account the type and distribution of the ICECAP-O scores (tariff or domain scores, as the response variable): (1) ordinary least squares [OLS] or censored least absolute deviation [CLAD] models were used to quantify the link between overall health status (EQ-5D-3L tariff score) or domains of health (EQ-5D-3L domain

scores) and overall capability (ICECAP-O tariff score as a continuous variable); (2) Multinomial Logistic [MNL] models to quantify the link between overall health status or domains of health and domains of capability (ICECAP-O domain scores as categorical variables).

ICECAP-O Tariff Score as a Continuous Variable

The OLS model is a commonly used model, particularly in the context of mapping studies.[39,72] On occasion, the OLS model provides a relatively good, if not superior, model performance compared with alternative models.[39,42,73] However, when data are semi-continuous, which has been shown to be a characteristic of EQ-5D-3L and ICECAP-O data, there is evidence that OLS may not be the best model.[41,74,75] In such circumstances, the CLAD model may be appropriate because it is robust in the presence of heteroscedasticity and non-normality, and allows a censoring (consistent with full capability at a tariff score of 1) at the upper end of the data distribution.[41,76]

ICECAP-O Domain Scores as Categorical Variables

The MNL model allows prediction of the domain scores of the ICECAP-O. This additional information could be used to describe the relationship between health status and capability at the domain level for both measures.[42,75,77] The MNL model assigns a probability to the likelihood of a person reporting a particular level score for each domain of the target measure, which is represented by the coefficient from the MNL model. The MNL model was estimated twice using a different number of Monte Carlo simulations (once with 1 simulation, once with 100 simulations) to assess the effect of running multiple simulations on model performance.[78] Monte Carlo simulation was preferred to other methods such as expected utility or most-likely probability methods,[79] and probabilistic mapping,[80] because it ensured that unbiased expected values were obtained.[78] This method previously performed relatively well in a mapping study with the ICECAP-O as the response variable.[42]

Model Specifications

Ten model specifications were assessed (see Table 1). Four of these model specifications (see
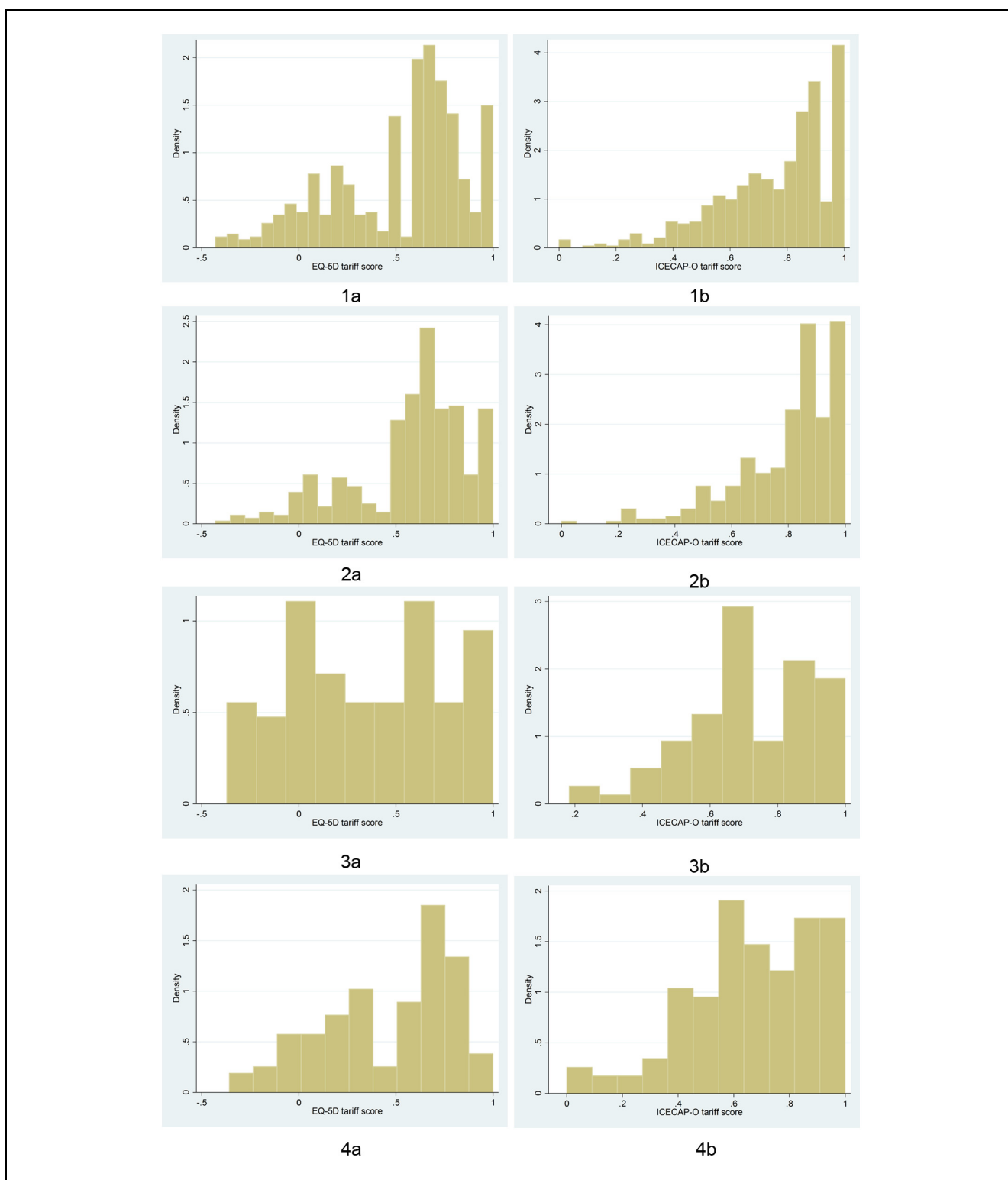
Figure 1   Distribution of the observed tariff scores for (a) EQ5D-3L and (b) ICECAP-O for 4 datasets: (1) MCOP, (2) AMOS, (3) BMH, and (4) CHOS.

Table 1 Selected Regression Model Specifications

Model	Response Variable(s)	Explanatory Variable(s)
Regression model: OLS and CLAD		
1	ICECAP-O tariff score	EQ-5D-3L tariff score, age, gender, care home
2	ICECAP-O tariff score	EQ-5D-3L tariff score
3	ICECAP-O tariff score	EQ-5D-3L domain scores, age, gender, care home
4	ICECAP-O tariff score	EQ-5D-3L domain scores
5	ICECAP-O tariff score	EQ-5D-3L items (continuous), age, gender, care home
6	ICECAP-O tariff score	EQ-5D-3L items (continuous)
7	ICECAP-O tariff score	EQ-5D-3L items (discrete), age, gender, care home
8	ICECAP-O tariff score	EQ-5D-3L items (discrete)
Regression model: MNL		
9[a]	ICECAP-O dimension scores	EQ-5D-3L tariff score
10[a]	ICECAP-O dimension scores	EQ-5D-3L items (discrete)

[a]Models were run: 1) as normal; 2) using multiple simulations (100 simulations), as recommended by Gray et al.[78] Both sets of results are reported in this paper.

Table 1, Models 1, 3, 5 and 7) add covariates (age, gender and care home [being a resident in a care home]) in line with published recommendations, with the aim of improving the statistical robustness of the model.[39,42,75,81]

Internal Validity

The "best" performing model specification was identified using tests for internal validity. "Best" was defined as the lowest absolute difference (AD) between the mean observed and predicted value; lowest mean absolute error (MAE); lowest root mean squared error (RMSE); and highest R^2 statistic. (Note, AD biases the results to preferring OLS over CLAD or MNL models but, given the properties and uses of the arithmetic mean, such a bias is beneficial when estimating and providing summary statistics describing the relationship between health and capability, which will include focus on the mean value.) Internal validity was also checked by comparing the results from the analysis of the estimation dataset (MCOP) with data from the 3 independent study samples (AMOS; BMH; CHOS). This analysis is classed as assessing internal, rather than external, validity because the 3 independent samples formed the MCOP sample.

RESULTS

Table 2 shows the demographic, screening tool (MMSE) and measure (ICECAP-O and EQ-5D-3L) scores information for the estimation dataset. The mean (standard deviation; SD) ICECAP-O and EQ-5D-3L scores in the estimation dataset (MCOP) were 0.76 (0.20) and 0.53 (0.34), respectively; lower than the mean score for the UK population over 75 y for ICECAP-O of 0.82[57] and the EQ-5D-3L of 0.73.[82] The highest mean (SD) ICECAP-O score across studies was for the AMOS study, 0.80 (0.18), which also had the relatively highest EQ-5D-3L score, 0.59 (0.30). The BMH study had a relatively higher mean ICECAP-O score than CHOS (0.71 v. 0.67); although, CHOS had a relatively higher mean EQ-5D-3L score than the BMH study (0.46 v. 0.35). As observed in Figures 1 and 2, only a small proportion of people across and within study samples had low EQ-5D-3L and ICECAP-O tariff scores (e.g., 33 [5.6%] people had an ICECAP-O tariff score <0.4).

Describing the Relationship between the EQ-5D-3L and ICECAP-O

Figure 2 shows the relationship between the EQ-5D-3L and ICECAP-O from the estimation dataset (MCOP). There was a positive relationship between the measures; although, this was not obvious on visual inspection of the scatter plot.

Quantifying the Relationship between EQ-5D-3L and ICECAP-O

Table 3 reports the results from the different model specifications of the estimation dataset (MCOP). Model 7 (OLS model with EQ-5D-3L items as discrete variables, including age, sex, and care home explanatory variables) produced the best model overall, with the lowest RMSE (0.1626) and

Table 2 Descriptive Statistics for the Combined Sample (MCOP) and Sample from Three Studies (AMOS; BMH; CHOS)

Characteristics	MCOP *n* = 584	AMOS *n* = 374	BMH *n* = 83	CHOS *n* = 127
Female, *n* (%)	363 (62.2%)	212 (56.7%)	51 (61.5%)	100 (78.7%)
Care home resident, *n* (%)	144 (24.7%)	1 (0.3%)	16 (19.3%)	127 (100%)
Cognitively impaired (MMSE score < 24), *n* (%)	182 (31.2%)	40 (10.7%)	45 (54.2 %)	97 (76%)
	Mean (SD, Range)			
Age	81.3 (7.1, 62–102)	79.8 (6.4, 70–99)	82.6 (6.53, 70–100)	85.12 (7.96, 62–102)
EQ-5D-3L tariff score	**0.53 (0.34, −0.429–1)**	**0.59 (0.30, −0.429–1)**	**0.35 (0.40, −0.371–1)**	**0.46 (0.34, −0.358–1)**
Mobility	*1.75 (0.54, 1–3)*	*1.69 (0.47, 1–3)*	*1.82 (0.77, 1–3)*	*1.88 (0.53, 1–3)*
Self-care	*1.65 (0.74, 1–3)*	*1.40 (0.58, 1–3)*	*2.17 (0.81, 1–3)*	*2.05 (0.75, 1–3)*
Usual activities	*1.86 (0.72, 1–3)*	*1.83 (0.69, 1–3)*	*1.70 (0.71, 1–3)*	*2.04 (0.79, 1–3)*
Pain/discomfort	*1.78 (0.61, 1–3)*	*1.86 (0.57, 1–3)*	*1.87 (0.58, 1–3)*	*1.46 (0.64, 1–3)*
Anxiety/depression	*1.48 (0.57, 1–3)*	*1.43 (0.54, 1–3)*	*1.61 (0.62, 1–3)*	*1.54 (0.60, 1–3)*
ICECAP-O tariff score	**0.76 (0.20, 0–1)**	**0.80 (0.18, 0–1)**	**0.71 (0.18, 0.18–1)**	**0.67 (0.23, 1–4)**
Attachment	*3.25 (0.88, 1–4)*	*3.35 (0.82, 1–4)*	*3.16 (0.93, 1–4)*	*3.03 (0.96, 1–4)*
Security	*2.87 (0.98, 1–4)*	*2.90 (0.90, 1–4)*	*2.80 (1.10, 1–4)*	*2.82 (1.12, 1–4)*
Role	*2.65 (0.98, 1–4)*	*2.85 (0.89, 1–4)*	*2.29 (1.03, 1–4)*	*2.30 (1.03, 1–4)*
Enjoyment	*2.72 (0.90, 1–4)*	*2.79 (0.87, 1–4)*	*2.54 (0.87, 1–4)*	*2.62 (0.96, 1–4)*
Control	*2.80 (0.91, 1–4)*	*3.04 (0.83, 1–4)*	*2.60 (0.81, 1–4)*	*2.21 (0.91, 1–4)*

MCOP, combined sample of 3 datasets for Medical Crises in Older People [program]; AMOS, dataset for Acute Medical Outcomes Study; BMH, dataset for Better Mental Health [study]; CHOS, dataset for Care Home Outcomes Study; SD, standard deviation.
MMSE, Mini Mental State Examination. The MMSE is a screening tool for cognitive impairment,[67,68] the score for which can be treated as a continuous (ranging from zero [cognitive impairment] to 30 [cognitive normality]) or discrete variable; the latter can be based on those groupings described by Folstein et al.[67] (cognitive impairment a MMSE score <24; cognitively normal a MMSE score ≥24).

highest R^2 (0.3532). The lowest MAE (0.1191) was produced by model 13 (CLAD model), which was the best CLAD model overall but with a higher RMSE (0.1654) and lower R^2 (0.3418) than model 7. The smallest AD was produced by each of the OLS models (models 1, 2, 3, 4, 5, and 6), which was intuitively correct, given that OLS is a linear mean model. The MNL models (models 9 and 10) performed worst overall across all statistics.

Table 4 reports the difference in results when the performance statistics from the internal validation assessment using the estimation dataset (MCOP) were compared head-to-head with the performance statistics from the same algorithms but when applied to the 3 study samples independently. The performance statistics and coefficients for the 20 regression models for the 3 independent study samples are provided in Appendix S3 and S4. This head-to-head comparison suggested that all model performance statistics improved in the AMOS sample compared with the estimation dataset (MCOP). These statistics worsened within the CHOS sample at a larger scale than any other sample (see Table 4);

across all models: MAE increased within the range of 0.0530 to 0.0715; RMSE increased within the range of 0.0613 to 0.0812; the R^2 statistic was lower
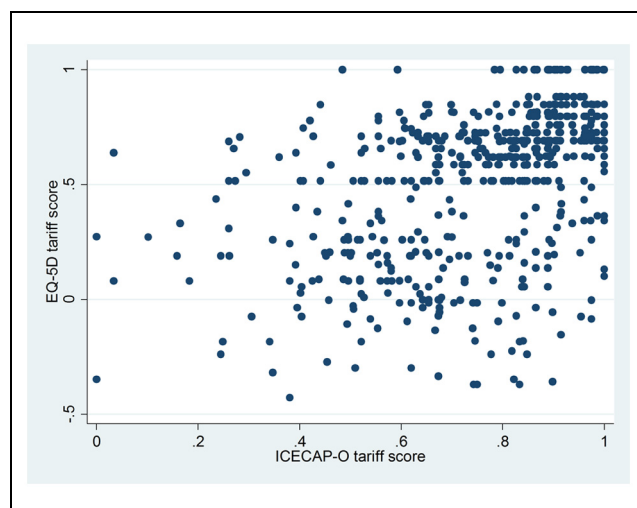

Figure 2 Scatter plot of the relationship between the EQ5D-3L and ICECAP-O tariff scores for the observed dataset from MCOP.

Table 3 Internal validation (MCOP)

Model	Spec'n[a]	Number	Observed Values			Estimated Values			Model Performance			
			ICECAP-O	Min	Max	ICECAP-O	Min	Max	AD	MAE	RMSE	R^2
Regression model: OLS												
1	1	584	0.7601	0	1	0.7601	0.4251	0.9342	0.0000	0.1279	0.1704	0.2787
2	2	584	0.7601	0	1	0.7601	0.4849	0.8969	0.0000	0.1334	0.1749	0.2365
3	3	584	0.7601	0	1	0.7601	0.3993	0.9246	0.0000	0.1220	0.1643	0.3336
4	4	584	0.7601	0	1	0.7601	0.4115	0.8952	0.0000	0.1233	0.1660	0.3165
5	5	584	0.7601	0	1	0.7601	0.4304	0.9370	0.0000	0.1212	0.1629	0.3456
6	6	584	0.7601	0	1	0.7601	0.4563	0.9169	0.0000	0.1224	0.1643	0.3305
<u>7</u>	<u>7</u>	<u>584</u>	<u>0.7601</u>	<u>0</u>	<u>1</u>	<u>0.7601</u>	<u>0.4484</u>	<u>0.9396</u>	***<u>0.0000</u>***	**0.1208**	***<u>0.1626</u>***	***<u>0.3532</u>***
8	8	584	0.7601	0	1	0.7601	0.4993	0.9281	0.0000	0.1221	0.1640	0.3387
Regression model: CLAD[b]												
9	1	584	0.7601	0	1	0.7831	0.4188	0.9719	0.0230	0.1253	0.1724	0.2777
10	2	584	0.7601	0	1	0.7873	0.4487	0.9557	0.0272	0.1291	0.1784	0.2365
11	3	584	0.7601	0	1	0.7766	0.3645	0.9635	**0.0165**	0.1204	0.1677	0.3241
12	4	584	0.7601	0	1	0.7847	0.4010	0.9456	0.0246	0.1212	0.1692	0.3118
13	5	584	0.7601	0	1	0.7824	0.4259	0.9730	0.0223	***0.1191***	**0.1654**	**0.3418**
14	6	584	0.7601	0	1	0.7800	0.4431	0.9560	0.0199	0.1200	0.1667	0.3261
15	7	584	0.7601	0	1	0.7838	0.3557	0.9810	0.0237	0.1200	0.1699	0.3229
16	8	584	0.7601	0	1	0.7813	0.4395	0.9624	0.0212	0.1206	0.1700	0.3138
Regression model: MNL												
17	9	584	0.7601	0	1	0.7642	0.1928	1	0.0041	0.1545	0.2080	0.1109
18	9	584*100	0.7601	0	1	0.7601	0	1	***0.0000***	0.1629	0.2141	0.0880
19	10	584	0.7601	0	1	0.7612	0	1	0.0011	**0.1467**	**0.1984**	**0.1892**
20	10	584*100	0.7601	0	1	0.7605	0	1	0.0004	0.1496	0.1998	0.1703

AD, absolute difference; MAE, mean absolute error; RMSE, root mean squared error.

[a]As defined in Table 1.

[b]CLAD regression model estimated in Stata using command as follows: clad var_i, ul(1) reps(200). Seed set value: 123456789.

Numbers in bold: performed best within statistic within model; numbers in italics: performed best within statistic across models; numbers underlined: best model across all model performance statistics.

with this difference in statistic value being in the range of 0.0771 to 0.2579 compared with the model's performance in the MCOP sample.

Quantifying the Relationship between Health and Capability

The best performing regression model was OLS model 7. Model 7 is a linear model and this may have implications when quantifying the relationship between different parts of the score distribution of the ICECAP-O. To account for this, Table 5 presents 2 performance statistics, MAE and RMSE, to show how the best performing OLS model (7) performs when quantifying the association for different parts of the ICECAP-O tariff score distribution.

The overall performance of OLS model 7 resulted in an MAE of 0.1208 and an RMSE of 0.1626. The higher RMSE compared with the MAE value was indicative of higher degrees of error between the observed and estimated values. Table 5 shows that the size of the error between the observed and estimated values was larger for values at the lower end of the tariff score and smaller when the observed values were closer to the mean value. For example, in the MCOP sample, when the observed value for the ICECAP-O score was in the range of 0.2 to 0.4, the MAE and RMSE were 0.3221 and 0.5162, respectively. Nearing the peak of the distribution, when the observed ICECAP-O tariff score was in the range of 0.6 to 0.8, the MAE and RMSE values were 0.0877 and 0.1136, respectively. This result was consistent among the 3 independent samples, where lower MAE and RMSE values were observed within the peak of the score distribution rather than the left-hand tail. A quantile-quantile (Q-Q) plot to further assess the potential bias induced by the OLS model 7 (particularly at lower level ICECAP-O tariff scores) is provided in Appendix S5.

Table 4 Regression Model Performance for Each Dataset

Regression Model Performance Statistic

Model	Spec'n[a]	Absolute Difference (AD)				Mean Absolute Error (MAE)				Root Mean Squared Error (RMSE)				R^2 Statistic			
		MCOP[b]	AMOS	BMH	CHOS	MCOP	AMOS	BMH	CHOS	MCOP	AMOS	BMH	CHOS	MCOP	AMOS	BMH	CHOS
Regression model: OLS																	
1	1	0.0000	0.0032	0.0020	0.0081	0.1279	−0.0206	0.0076	0.0559	0.1704	−0.0283	0.0115	0.0653	0.2787	0.1249	−0.1439	−0.2466
2	2	0.0000	0.0243	0.0031	0.0737	0.1334	−0.0199	0.0012	0.0579	0.1749	−0.0306	0.0010	0.0707	0.2365	0.1525	−0.0892	−0.2041
3	3	0.0000	0.0028	0.0047	0.0053	0.1220	−0.0198	0.0053	0.0549	0.1643	−0.0264	0.0055	0.0696	0.3336	0.0996	−0.0953	−0.2579
4	4	0.0000	0.0132	0.0030	0.0411	0.1233	−0.0186	0.0004	0.0547	0.1660	−0.0265	−0.0018	0.0692	0.3165	0.0985	−0.0579	−0.2339
5	5	0.0000	0.0028	0.0047	0.0052	0.1212	−0.0195	0.0064	0.0534	0.1629	−0.0259	0.0069	0.0675	0.3456	0.0917	−0.1041	−0.2529
6	6	0.0000	0.0129	0.0011	0.0387	0.1224	−0.0186	0.0024	0.0530	0.1643	−0.0258	0.0006	0.0665	0.3305	0.0895	−0.0752	−0.2301
7	7	0.0000	0.0037	0.0101	0.0042	0.1208	−0.0199	0.0086	0.0531	0.1626	−0.0256	0.0144	0.0707	0.3532	0.0884	−0.1236	−0.2503
8	8	0.0000	0.0136	0.0062	0.0351	0.1221	−0.0192	0.0045	0.0534	0.1640	−0.0255	0.0075	0.0697	0.3387	0.0875	−0.0966	−0.2318
Regression model: CLAD																	
9	1	0.0230	−0.0021	−0.0045	0.0090	0.1253	−0.0230	0.0120	0.0598	0.1724	−0.0306	0.0141	0.0685	0.2777	0.1256	−0.1394	−0.2455
10	2	0.0272	−0.0203	−0.0145	0.0693	0.1291	−0.0261	0.0079	0.0715	0.1784	−0.0380	0.0064	0.0812	0.2365	0.1525	−0.0892	−0.2042
11	3	0.0165	−0.0012	−0.0018	0.0044	0.1204	−0.0219	0.0061	0.0604	0.1677	−0.0295	0.0056	0.0753	0.3241	0.1003	−0.0748	−0.2555
12	4	0.0246	−0.0113	−0.0090	0.0393	0.1212	−0.0217	0.0023	0.0624	0.1692	−0.0312	−0.0005	0.0770	0.3118	0.1079	−0.0617	−0.2349
13	5	0.0223	−0.0009	−0.0012	0.0036	0.1191	−0.0211	0.0062	0.0580	0.1654	−0.0271	0.0067	0.0702	0.3418	0.0902	−0.0864	−0.2524
14	6	0.0199	−0.0097	−0.0139	0.0375	0.1200	−0.0213	0.0026	0.0608	0.1667	−0.0291	0.0010	0.0727	0.3261	0.0951	−0.0626	−0.2339
15	7	0.0237	−0.0039	−0.0061	0.0155	0.1200	−0.0222	0.0091	0.0594	0.1699	−0.0293	0.0130	0.0793	0.3229	0.0947	−0.0863	−0.2309
16	8	0.0212	−0.0114	−0.0081	0.0388	0.1206	−0.0215	0.0035	0.0610	0.1700	−0.0301	0.0054	0.0799	0.3138	0.0968	−0.0657	−0.2200
Regression model: MNL																	
17	9	0.0041	0.0130	0.0060	0.0646	0.1545	−0.0105	0.0137	0.0637	0.2080	−0.0188	0.0195	0.0645	0.1109	−0.0085	−0.0913	−0.0980
18	10	0.0000	0.0232	0.0038	0.0752	0.1629	−0.0201	0.0053	0.0535	0.2141	−0.0272	0.0032	0.0615	0.0880	0.0451	−0.0220	−0.0771
19	9	0.0011	0.0060	0.0325	0.0095	0.1467	−0.0181	0.0022	0.0535	0.1984	−0.0252	0.0139	0.0622	0.1892	0.0236	−0.1170	−0.1042
20	10	0.0004	0.0126	0.0162	0.0302	0.1496	−0.0215	0.0063	0.0554	0.1998	−0.0281	0.0018	0.0613	0.1703	0.0471	−0.0653	−0.1164

[a]As defined in Table 1.
[b]The MCOP sample is defined as the primary sample for this analysis; underlined values are the MCOP baseline statistics.

Table 5 MAE and RMSE of Estimated V. Observed Scores by ICECAP-O Tariff Score Groups for Best Performing Model (OLS Model 7)

| | Combined Dataset | | | | Individual Datasets | | | | | | | | | | | |
| | MCOP | | | | AMOS | | | | BMH | | | | CHOS | | | |
ICECAP-O Score	Number	%	MAE	RMSE	Number	%	MAE	RMSE	Number	%	MAE	RMSE	Number	%	MAE	RMSE
0 < 0.2	8	1.4	0.485	N/A	2	0.5	0.485	N/A	1	1.2	0.365	N/A	5	3.9	0.510	N/A
0.2 < 0.4	25	4.3	0.322	0.516	11	2.9	0.336	N/A	4	4.8	0.377	N/A	10	7.9	0.285	N/A
0.4 < 0.6	85	14.5	0.155	0.200	39	10.4	0.172	0.237	15	18.1	0.123	0.578	31	24.4	0.150	0.255
0.6 < 0.8	157	26.9	0.080	0.114	86	23.0	0.083	0.110	30	36.1	0.101	0.166	41	32.3	0.092	0.140
0.8 < 1	309	52.9	0.102	0.137	236	63.1	0.082	0.109	33	39.8	0.121	0.184	40	31.5	0.207	0.295
Full index	584	100	0.121	0.163	374	100	0.101	0.137	83	100	0.129	0.177	127	100	0.174	0.233

AD, absolute difference; MAE, mean absolute error; RMSE, root mean squared error; N/A, not applicable (in this instance due to the small sample size).

The coefficients for the explanatory variables and intercepts estimated by the best performing OLS model 7 are presented in Table 6. Using these coefficients, the following equation represents the best performing regression model and the best estimate of a quantified relationship between health and capability:

$$Overall\ ICECAP\text{-}O\ tariff\ score=$$
$$1.044 - 0.002 * age$$
$$+ 0.009 * female - 0.051 * carehome$$
$$- 0.016 * s\_mo + 0.010 * e\_mo$$
$$- 0.081 * s\_sc - 0.159 * e\_sc$$
$$- 0.042 * s\_ua - 0.106 * e\_ua$$
$$- 0.003 * s\_pd + 0.032 * e\_pd$$
$$- 0.089 * s\_ad - 0.083 * e\_ad$$

These coefficients can each be interpreted as an associative (not causal) relationship between health and capability when we have accounted for the effects of all other variables in the model. Not all the estimated associations were statistically significant (assuming statistical significance is defined at a 5% threshold level; $P < 0.05$); therefore, for descriptive purposes, only statistically significant relationships are now described. Moving from "no problem" with self-care to: "some problem" is associated with a decrease in capability of 0.081 ($P = 0.000$); and "extreme problem" is associated with a decrease in capability of 0.159 ($P = 0.000$). Moving from "no problem" with usual activities to "extreme problem" is associated with a decrease in capability of 0.106 ($P = 0.000$). Moving from "no problem" with anxiety/depression to: "some problem" is

associated with a decrease in capability of 0.089 ($P = 0.000$); and "extreme problem" is associated with a decrease in capability of 0.083 ($P = 0.030$). Living in a care home was also statistically significantly associated with a decrease in capability of 0.051 relative to living in the community ($P = 0.007$).

DISCUSSION

Regression methods, synonymous with those applied in mapping studies, were used in this study because they provided the relevant basis to identify the extent of the quantifiable link between 2 measures and their underlying constructs.[39–44] The results of this study suggest it was possible to quantify a relationship between the EQ-5D-3L and ICECAP-O (albeit, with large errors around the point estimate coefficients). Results from the best performing regression model (OLS model 7) suggested that capability did have a statistically significant relationship with some domains of health (self-care; usual-activities; anxiety/depression), but not all domains of health (mobility; pain/discomfort) at the 5% significance level; the small number of very low score observations limits the extent to which a statistically significant result could be detected, which should be taken into account when interpreting these results. This result could suggest that ICECAP-O does not include the domains of capability with which mobility or pain/discomfort would have a relationship, and so would be an insensitive measure for assessing change in capability for interventions focused on improving these aspects of health. Alternatively, it may suggest that a change in mobility or pain/discomfort is generally not

Table 6 Best Performing Regression Model Quantifying the Link between ICECAP-O Capability Index and the EQ-5D-3L Item Scores and Covariates (OLS Model 7)

Variable	Description	Equation Notation	Coefficient	Standard Error (SE)	95% Confidence Interval	P Value
Age	Age of the patient	*age*	−0.002	0.001	−0.004 to 0.000	0.097
Female	Gender	*female*	0.009	0.014	−0.019 to 0.037	0.513
Care home	Care home resident	*carehome*	−0.051	0.019	−0.088 to −0.014	0.007
Mobility	EQ-5D-3L domain					
Some problem	*Level 2*	*s_mo*	−0.016	0.018	−0.052 to 0.021	0.398
Extreme problem	*Level 3*	*e_mo*	0.010	0.038	−0.065 to 0.085	0.788
Self-care	EQ-5D-3L domain					
Some problem	*Level 2*	*s_sc*	−0.081	0.018	−0.116 to −0.047	0.000
Extreme problem	*Level 3*	*e_sc*	−0.159	0.037	−0.208 to −0.109	0.000
Usual activities	EQ-5D-3L domain					
Some problem	*Level 2*	*s_ua*	−0.042	0.017	−0.076 to −0.007	0.017
Extreme problem	*Level 3*	*e_ua*	−0.106	0.023	−0.151 to −0.062	0.000
Pain/discomfort	EQ-5D-3L domain					
Some problem	*Level 2*	*s_pd*	−0.003	0.017	−0.037 to 0.031	0.883
Extreme problem	*Level 3*	*e_pd*	0.032	0.028	−0.088 to 0.024	0.265
Anxiety/depression	EQ-5D-3L domain					
Some problem	*Level 2*	*s_ad*	−0.089	0.015	−0.118 to −0.060	0.000
Extreme problem	*Level 3*	*e_ad*	−0.083	0.038	−0.157 to −0.008	0.030
Constant	Constant term	N/A	1.044	0.083	0.881 to 1.206	0.000

statistically significantly associated with a change in capability. This would mean that the generalized conceptual idea that health is a conversion factor for capability is not true for all domains of health.

Findings from this study add further support that the 2 constructs of health status and capability—when quantified using EQ-5D-3L and ICECAP-O, respectively—are complements rather than direct substitutes for each other. Keeley et al.[83] have also explored the link between capability and health using a different capability-based measure, suggesting ICECAP-A and EQ-5D-3L were measuring 2 different constructs, producing different but complementary information, a result which was further supported by Engel et al.[84] (comparing ICECAP-A and EQ-5D-5L) and Davis et al.[36] (comparing ICECAP-O and EQ-5D-3L). This study indicates that it is not possible to produce a robust mapping algorithm based on the conceptual and design differences of the ICECAP-O and EQ-5D-3L.

Together with previous studies,[36,83,84] this study questions whether the measures conceptually overlap sufficiently in their descriptive systems to support the face validity of using a mapping algorithm.[41] The 2 measures operate on different numerical and conceptual scales. For descriptive purposes, assume that a value of zero is equivalent across both

scales (i.e., a state equivalent to "dead" is the same as "no capability"; although, conceptually, "no capability" might not be the same as "dead"). It is logical to assume that, as health declines, so does capability (this hypothesis is supported by the results in this study) up to the point where a value of zero is reported across both measures (i.e., a state equivalent to dead is equal to no capability). This assumption is conceptually possible but practically could never happen, because zero is not an achievable EQ-5D-3L tariff score. However, when using the EQ-5D-3L, health can decline into negative values (or "states worse than dead") but there are no negative values for the ICECAP-O (i.e., there are no assumed negative capability states). In this case, it is quite possible for a person to be in "a state worse than dead" (e.g., −0.2) and have a positive value of capability (e.g., 0.2), while still assuming that the values of zero across both scales are equivalent; conceptually this is illogical and adds further concern to producing and using a mapping algorithm.

The methods used to generate the available tariff scores for the 2 measures, ICECAP-O and EQ-5D-3l, best-worst scaling[55] and time–trade-off,[65] respectively, did not account for the impact of time or anchoring in an equivalent way. Brazier et al.[85] provide a useful discussion about these issues when

eliciting preference-based scales, which is beyond the scope of this paper. Furthermore, the feasible range of the scores and data scaling issues for each outcome cause measurement issues when using regression analysis. Practically, it is still feasible to perform the regression analysis but the measurement issues mean that, when estimating a mapping function from a larger scale to a smaller scale (i.e., -0.594 to one [EQ-5D-3L] or zero to one [ICECAP-O]), there will be a corresponding change in the scale of the coefficients but no change in statistical significance. Therefore, in this instance the estimated coefficients may be smaller, which will have implications when quantifying and describing the relationship between health and capability. This means that mapping between measures and scales would not appropriately account for the conceptual differences in terms of what the scales and scores mean and their subsequent application in economic evaluations.

This study suggested there was a substantial difference in the mean EQ-5D-3L score but a marginal difference in the mean ICECAP-O score for the MCOP dataset when compared with the general population (0.53 v. 0.73 for EQ-5D-3L; 0.76 v. 0.82 for ICECAP-O). After simply rescaling to account for the EQ-5D-3L tariff score scale (1/1.594 = 0.627), the absolute difference in ICECAP-O and EQ-5D-3L scores could be described as a factor of 2 (0.06 v. 0.13). This observed difference in scores relative to the general population is logical and feasible. The empirical literature supports that people tend to adapt to their state of being (such as health),[16,86,87] which indicates that although in a lower health state, a person's capability may start to move back towards "normal" as they adapt to the health state (where "normal" could be defined as the general population scores in this instance), when compared with the general population. For example, someone in a wheelchair might have the "ability to achieve independence" (an ICECAP-O domain), even though their mobility is severely impaired (an EQ-5D-3L domain). The impact of if, and how, people adapt is an external factor that cannot easily be understood or accounted for when using mapping methods; this is likely to restrict the generalizability of any estimated algorithm measuring the relationship between health and capability-based measures.

The relatively low levels of health status and capability in the MCOP dataset and the assumptions of the regression models used (e.g., OLS) will also influence the robustness of the observed quantified relationship between health and capability.

When assessing the validity of the quantified relationship using the 3 study samples independently, performance of the models improved when using the AMOS dataset (patients with relatively better quality of life than the overall MCOP sample) compared with the MCOP dataset. Model performance was generally not as good in the BMH and CHOS datasets (patients with poorer quality of life than the overall MCOP sample). This result is echoed in the results assessing the performance of the best performing model for estimating different parts of the ICECAP-O tariff score distribution, whereby performance worsened at the lower end of the tariff score distribution; this suggests that the quantified relationship may not well explain the relationship between poorer health and lower levels of capability.

Limitations and Recommendations for Future Work

The estimation dataset used in this study was relatively small compared with some previous mapping studies.[80,81] However, "successful mapping" (in terms of better performance statistics) has been conducted on smaller sample sizes.[39,42,75] Furthermore, there were no baseline ICECAP-O data in the available dataset, so it was not possible to assess the potential sensitivity of the ICECAP-O to change over time.

In terms of validation, the preferred method is to use an external validation sample but this was not feasible in this study.[41] Validation of the regression specification selected in this study was limited to internal validation using 3 sub-groups defined a priori. Other approaches to internal validation, such as K-fold,[88–91] could have been explored but the impact of using this approach is a topic for future research. During internal validation, the poorest performance statistics were observed for the CHOS sample. This could be attributed to: (i) the use of proxy and self-responses within this study, when previous studies have shown a discrepancy between proxy and self-response;[92,93] (ii) the nature of care homes and their residents (for example, poor health and no other option [in relation to informal or formal carers enabling community living], forcing a need to live in a care home) may change the relationship between health and capability compared with their community-dwelling counterparts for which the selected model specification may be better suited.

The EQ-5D-3L is a relatively specific measure of generic health status. There are more comprehensive (e.g., SF-36[94]) and condition-specific health status (e.g., QLQ-C30 for cancer[95]) measures that could have been compared with capability. A newer alternative to the EQ-5D-3L, the EQ-5D-5L,[96,97] which has 5 rather than 3 levels, is also recommended by NICE.[5] It was not possible to use EQ-5D-5L in this study, as it was not ready for use during the time period of the MCOP program's studies. People may respond differently to the EQ-5D-5L,[98] leading to a redistribution of responses and therefore a change in how health is quantified and described. Using the EQ-5D-5L could affect the quantified relationship between capability and health status. It is difficult to hypothesize the effect this redistribution might have on the estimated relationship and this should be a focus for future research. Other capability-based measures could also be considered to further explore the relationship between health status and capability, such as ASCOT,[99] OCAP-18,[100] or OxCAP-MH[101] measures. Consistent with Makai et al.[102], future studies should assess the relationship between other aspects of health and capability, particularly if these constructs are to be used as objective endpoints by which the effectiveness of interventions will be judged within an economic evaluation.

A general limitation of the approach used in this study was using OLS and CLAD models for complex score distributions (for example, non-normally distributed data with multiple peaks). The linear aspect of these models potentially limited the extent to which the estimated relationship captured the impact over different points of the distribution of the scores for the target measure within a defined patient group. Additional model specifications, such as mixture models to assess different aspects of the distribution,[75,77] could have been explored but these models would still not compensate for the inherent limitations in the available dataset. The small number of patients with very low ICECAP-O scores (<0.4; $n = 33$) in the MCOP dataset meant that there may be too few observations to detect a statistically significant association between lower levels of health and capability. The small number of low-score observations limits the production of a robust algorithm at the lower levels of health and capability using regression analysis, which is a limitation for this study and other studies where it is difficult to recruit patients with these low scores (e.g., a low score is associated with poor health, and poor health restricts a person's ability to take part in

research). This limits the generalizability of the quantified relationship to "all" older people (e.g., different co-morbidities and care consumption, including medications) and other countries (e.g., those with different health and social care systems, which may affect health and capability).

It was not possible to make causal inferences about the relationship between the 2 constructs. Establishing causal inferences would require fitting regression models that contain different combinations of 1 to 4 domains and interactions between covariates (such as being in a care home and type of care home) in a larger sample size with a wider distribution of outcome measure scores (i.e., more lower level scores for both measures) than that available in this study. Therefore, exploring causal inferences could be the focus of future research but would require a more suitable dataset.

CONCLUSION

A statistically significant association with capability (measured using ICECAP-O) was identified for 3 (self-care, usual-activities, and anxiety/depression) of the 5 EQ-5D-3L domains. Although health status was found to be positively and directly associated with capability, the strength of the association suggested that it is not appropriate to use a mapping algorithm to provide a link between the EQ-5D-3L and ICECAP-O. This study demonstrated how the relationship between health and capability can be assessed using regression-based methods and adds further support to previously published studies that a measure of capability, in this case, the ICECAP-O, is providing complementary information rather than acting as a direct substitute to a measure of health status.

involved in the entire MCOP programme of studies. The authors would also like to acknowledge the wider MCOP study group which included John Gladman, Simon Conroy, Rowan Harwood, Anthony Avery, Sarah Lewis, Davina Porock, Rob Jones, Pip Logan, Justine Schneider, Jane Dyas, Judi Edmans, Adam Gordon, Sarah Goldberg, Vladislav Berdunov, Lukasz Tanajewski, Georgios Gkountouras, Lucy Bradshaw and Bella Robbins.

REFERENCES

1. Torrance GW, Feeny D. Utilities and quality-adjusted life years. Int J Technol Assess Health Care. 1989;5:559–75.

2. Vergel YB, Sculpher M. Quality-adjusted life years. Prac Neurol. 2008;8:175–82.

3. Karimi M, Brazier J. Health, health-related quality of life, and quality of life: What is the difference? Pharmacoeconomics. 2016; 34:645–9.

4. Wisløff T, Hagen G, Hamidi V, Movik E, Klemp M, Olsen JA. Estimating QALY gains in applied studies: a review of cost-utility analyses published in 2010. Pharmacoeconomics. 2014;32:367–75.

5. NICE. Guide to the methods of technology appraisal. National Institute for Health and Care Excellence (NICE). 2013.

6. EuroQol group. EuroQol-a new facility for the measurement of health-related quality of life. Health Policy. 1990;16:199–208.

7. Coast J. Is economic evaluation in touch with society's health values? BMJ. 2004;329:1233–6.

8. Coast J, Smith RD, Lorgelly P. Welfarism, extra-welfarism and capability: The spread of ideas in health economics. Soc Sci Med. 2008;67:1190–8.

9. Oliver A, Healey A, Donaldson C. Choosing the method to match the perspective: economic assessment and its implications for health-services efficiency. Lancet. 2002;359:1771–4.

10. Cookson R. QALYs and the capability approach. Health Econ. 2005;14:817–29.

11. Culyer AJ. The normative economics of health care finance and provision. Oxford Rev Econ Policy. 1989;5:34–58.

12. Payne K, McAllister M, Davies LM. Valuing the economic benefits of complex interventions: when maximising health is not sufficient. Health Econ. 2013;22:258–71.

13. Ryan M, Netten A, Skåtun D, Smith P. Using discrete choice experiments to estimate a preference-based measure of outcome—an application to social care for older people. J Health Econ. 2006;25:927–44.

14. Netten A, Burge P, Malley J, et al. Outcomes of social care for adults: developing a preference-weighted measure. Health Tech Asses. 2012;16:1–166.

15. Sen A. The idea of justice. Cambridge, MA: Harvard University Press; 2011.

16. Sen A. Commodities and Capabilities. 19th ed. Oxford: Oxford University Press; 2013.

17. Sen A. Capability and well-being. In: Nussbaum M, Sen A, eds. The quality of life. Oxford: Clarendon Press; 1993.

18. Robeyns I. An unworkable idea or a promising alternative?: Sen's capability approach re-examined. In: Center for Economic

Studies (CES). Discussion Paper Series (DPS) 0030. Katholleke Universiteit, Leuven; 2000.

19. Robeyns I. The capability approach: a theoretical survey. J Hum Dev. 2005;6:93–117.

20. Sen A. Inequality reexamined. Oxford: Oxford University Press; 1992.

21. Lorgelly PK. Choice of outcome measure in an economic evaluation: A potential role for the capability approach. Pharmacoeconomics. 2015;33:849–55.

22. Coast J, Kinghorn P, Mitchell P. The development of capability measures in health economics: opportunities, challenges and progress. Patient. 2015;8:119–26.

23. Flynn TN, Huynh E, Peters TJ, et al. Scoring the ICECAP-A Capability Instrument. Estimation of a UK General Population Tariff. Health Econ. 2015;24:258–69.

24. Mitchell P, Roberts T, Barton P, Coast J. Assessing sufficient capability: A new approach to economic evaluation. Soc Sci Med. 2015;139:71–9.

25. ICECAP measures website. ICECAP-O questionnaire. 2017; Available from: URL: http://www.birmingham.ac.uk/Documents/ college-mds/haps/projects/icecap/questionnaires/ ICECAP-Oquestionnaire.pdf.

26. Coast J, Flynn TN, Natarajan L, et al. Valuing the ICECAP capability index for older people. Soc Sci Med. 2008;67:874–82.

27. Al-Janabi H, Coast J. ICECAP-A: Developing a measure of adult's capabilities. Patient Report Outcomes. 2009;42:7–8.

28. NICE. The social care guidance manual. National Institute for Health and Care Excellence; 2013.

29. Makai P, Looman W, Adang E, Melis R, Stolk E, Fabbricotti I. 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;16:437–50.

30. Mitchell PM, Venkatapuram S, Richardson J, Iezzi A, Coast J. Are quality-adjusted life years a good proxy measure of individual capabilities? Pharmacoeconomics. 2017;35:637–46.

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

32. Sen A. Capabilities, lists, and public reason: continuing the conversation. Feminist Econ. 2004;10:77–80.

33. Nussbaum MC. Women and Human Development: The capabilities approach. Cambridge: Cambridge University Press; 2000.

34. Nussbaum MC. Symposium on Amartya Sen's philosophy: 5 Adaptive preferences and women's options. Econ Phil. 2001;17: 67–88.

35. Grewal I, Lewis J, Flynn T, Brown J, Bond J, Coast J. Developing attributes for a generic quality of life measure for older people: Preferences or capabilities? Soc Sci Med. 2006;62: 1891–901.

36. 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.

37. Coast J, Peters TJ, Natarajan L, Sproston K, Flynn T. An assessment of the construct validity of the descriptive system for the ICECAP capability measure for older people. Qual Life Res. 2008;17:967–76.

38. Couzner L, Ratcliffe J, Crotty M. The relationship between quality of life, health and care transition: an empirical comparison in an older post-acute population. Health Qual Life Outcomes. 2012;10:69–78.

39. Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. Eur J Health Econ. 2010;11:215–25.

40. Dakin H. Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. Health Qual Life Outcomes. 2013;11:1–6.

41. Longworth L, Rowen D. NICE DSU Technical Decision Support Document 10: The use of mapping methods to estimate health state utility values. In: ScHARR UoS, editor. Report by the Decision Support Unit (DSU). Sheffield 2011. p 1–31.

42. Mitchell PM, Roberts TE, Barton PM, Pollard BS, Coast J. Predicting the ICECAP-O Capability Index from the WOMAC Osteoarthritis Index: Is Mapping onto Capability from Condition-Specific Health Status Questionnaires Feasible? Med Decis Making. 2013;33:547–57.

43. Petrou S, Rivero-Arias O, Dakin H, et al. Preferred reporting items for studies mapping onto preference-based outcome measures: The MAPS statement. Health Qual Life Outcomes. 2015;33:985–91.

44. Wailoo AJ, Hernandez-Alava M, Manca A, et al. Mapping to estimate health-state utility from non–preference-based outcome measures: An ISPOR good practices for Outcomes Research Task Force Report. Value Health. 2017;20:18–27.

45. Edmans J, Bradshaw L, Gladman JR, et al. The Identification of Seniors at Risk (ISAR) score to predict clinical outcomes and health service costs in older people discharged from UK acute medical units. Age Ageing. 2013;42:747–53.

46. Franklin M, Berdunov V, Edmans J, et al. Identifying patient-level health and social care costs for older adults discharged from acute medical units in England. Age Ageing. 2014;43:703–7.

47. Bradshaw LE, Goldberg SE, Lewis SA, et al. Six-month outcomes following an emergency hospital admission for older adults with co-morbid mental health problems indicate complexity of care needs. Age Ageing. 2013;42:582–8.

48. Goldberg SE, Whittamore KH, Harwood RH, Bradshaw LE, Gladman JR, Jones RG. The prevalence of mental health problems among older adults admitted as an emergency to a general hospital. Age Ageing. 2012;41:80–6.

49. Hodkinson H. Evaluation of a mental test score for assessment of mental impairment in the elderly. Age Ageing. 1972;1:233–8.

50. Almeida OP, Almeida SA. Short versions of the geriatric depression scale: a study of their validity for the diagnosis of a major depressive episode according to ICD- 10 and DSM- IV. Int J Geriatr Psychiatry. 1999;14:858–65.

51. Spitzer RL, Williams JB, Kroenke K, et al. Utility of a new procedure for diagnosing mental disorders in primary care: the PRIME-MD 1000 study. JAMA. 1994;272:1749–56.

52. Ewing JA. Detecting alcoholism: the CAGE questionnaire. JAMA. 1984;252:1905–7.

53. Gordon AL, Franklin M, Bradshaw L, Logan P, Elliott R, Gladman JR. Health status of UK care home residents: a cohort study. Age Ageing. 2013;43:97–103.

54. TPP. SystmOne. The Phoenix Partnership; 2016; Available from: URL: https://www.tpp-uk.com/products/systmone. [Accessed 14 November, 2016].

55. Flynn TN, Louviere JJ, Peters TJ, Coast J. Best–worst scaling: what it can do for health care research and how to do it. J Health Econ. 2007;26:171–89.

56. Couzner L, Ratcliffe J, Lester L, Flynn T, Crotty M. Measuring and valuing quality of life for public health research: application of the ICECAP-O capability index in the Australian general population. Int J Public Health. 2013;58:367–76.

57. Flynn T, Chan P, Coast J, Peters T. Assessing Quality of Life among British Older People Using the ICEPOP CAPability (ICECAP-O) Measure. Appl Health Econ Health Policy. 2011;9: 317–29.

58. Ratcliffe J, Laver K, Couzner L, Quinn S, Corotty M. An assessment of the construct validity of the icecap-o index of capability in Australian national transition care and clinical rehabilitation programmes. In: Flinders Centre for Clinical Change and Health Care Research FU (ed). Flinders University (online) 2011. p 1–26.

59. Makai P, Beckebans F, van Exel J, Brouwer WB. Quality of life of nursing home residents with dementia: validation of the German version of the ICECAP-O. PLoS One. 2014;9:e92016 (1–10).

60. Makai P, Brouwer WB, Koopmanschap MA, Nieboer AP. Capabilities and quality of life in Dutch psycho-geriatric nursing homes: an exploratory study using a proxy version of the ICECAP-O. Qual Life Res. 2012;21:801–12.

61. Makai P, Koopmanschap MA, Brouwer WB, Nieboer AA. A validation of the ICECAP-O in a population of post-hospitalized older people in the Netherlands. Health Qual Life Outcomes. 2013;11:1–11.

62. Horwood J, Sutton E, Coast J. Evaluating the Face Validity of the ICECAP-O Capabilities Measure: A ''Think Aloud'' Study with Hip and Knee Arthroplasty Patients. Appl Res Qual Life. 2013:1–16.

63. EuroQol group website. EQ-5D-3L User Guide, 2017; Available from: URL: https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-3L_UserGuide_2015.pdf. [Accessed 31 July, 2017].

64. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997;35:1095–108.

65. Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: Results from a general population study. Health Econ. 1996;5:141–54.

66. Haywood K, Garratt A, Fitzpatrick R. Quality of life in older people: a structured review of generic self-assessed health instruments. Qual Life Res. 2005;14:1651–68.

67. Folstein MF, Folstein SE, McHugh PR. ''Mini-mental state'': a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12:189–98.

68. Cockrell J, Folstein M. Mini-Mental State Examination (MMSE). Psychopharmacol Bull. 1988;24:689–92.

69. Graham JE, Rockwood K, Beattie BL, et al. Prevalence and severity of cognitive impairment with and without dementia in an elderly population. Lancet. 1997;349:1793–6.

70. Carlo A, Baldereschi M, Amaducci L, et al. Cognitive impairment without dementia in older people: prevalence, vascular risk factors, impact on disability. The Italian Longitudinal Study on Aging. J Am Geriatrics Soc. 2000;48:775–82.

71. StataCorp. Stata Statistical Software: Release 11. College Station, TX: StataCorp LP; 2009.

72. Fryback DG, Dunham NC, Palta M, et al. US norms for six generic health-related quality-of-life indexes from the National Health Measurement study. Med Care. 2007;45:1162–70.

73. Dakin H, Petrou S, Haggard M, Benge S, Williamson I. Mapping analyses to estimate health utilities based on responses to the OM8-30 otitis media questionnaire. Qual Life Res. 2010;19:65–80.

74. Kennedy P. A Guide to Econometrics. Sixth ed. Cambridge, MA: MIT Press; 2008.

75. Hernández Alava M, Wailoo AJ, Ara R. Tails from the peak district: adjusted limited dependent variable mixture models of EQ-5D questionnaire health state utility values. Value Health. 2012;15:550–61.

76. Rowen D, Brazier J, Roberts J. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? Health Qual Life Outcomes. 2009;7:1–9.

77. Huang I, Frangakis C, Atkinson MJ, et al. Addressing ceiling effects in health status measures: a comparison of techniques applied to measures for people with HIV disease. Health Serv Res. 2008;43:327–39.

78. Gray AM, Rivero-Arias O, Clarke PM. Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. Med Decis Making. 2006;26:18–29.

79. Tsuchiya A, Brazier J, McColl E, Parkin D. Deriving preference-based single indices from non-preference based condition-specific instruments: Converting AQLQ into EQ5D indices. In: ScHARR UoS, (ed). HEDS Discussion Paper 02/01. University of Sheffield; 2002. p 1–45.

80. Le QA, Doctor JN. Probabilistic mapping of descriptive health status responses onto health state utilities using Bayesian networks: an empirical analysis converting SF-12 into EQ-5D utility index in a national US sample. Med Care. 2011;49:451–60.

81. Kaambwa B, Billingham L, Bryan S. Mapping utility scores from the Barthel index. Eur J Health Econ. 2013;14:231–41.

82. Kind P, Hardman G, Macran S. UK population norms for EQ-5D. In: York Uo, editor. Discussion Paper 172. University of York: Centre for Health Economics (CHE); 1999. p 1–98.

83. Keeley T, Coast J, Nicholls E, Foster N, Jowett S, Al-Janabi H. An analysis of the complementarity of ICECAP-A and EQ-5D-3 L in an adult population of patients with knee pain. Health Qual Life Outcomes. 2016;14:36–41.

84. Engel L, Mortimer D, Bryan S, Lear SA, Whitehurst DG. An investigation of the overlap between the ICECAP-A and five preference-based health-related quality of life instruments. PharmacoEconomics. 2017;35:741–53.

85. Brazier J, Rowen D, Yang Y, Tsuchiya A. Comparison of health state utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale. Eur J Health Econ. 2012;13:575–87.

86. Ubel PA, Loewenstein G, Jepson C. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. Qual Life Res. 2003;12:599–607.

87. Burchardt T. Agency goals, adaptation and capability sets. J Hum Dev Capabilities. 2009;10:3–19.

88. Anguita D, Ghelardoni L, Ghio A, Oneto L, Ridella S, editors. The 'K' in K-fold Cross Validation. Bruges: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning; 2012.

89. Blum A, Kalai A, Langford J, (eds). Beating the hold-out: Bounds for k-fold and progressive cross-validation. Proceedings of the 12th Annual Conference on Computational Learning Theory; New York: ACM; 1999.

90. Efron B, Tibshirani RJ. An introduction to the bootstrap. Boca Raton, FL: CRC press; 1994.

91. Kohavi R (ed). A study of cross-validation and bootstrap for accuracy estimation and model selection. Stanford, CA: Stanford University Press; 1995.

92. Tamim H, McCusker J, Dendukuri N. Proxy reporting of quality of life using the EQ-5D. Med Care. 2002;40:1186–95.

93. McPhail S, Beller E, Haines T. Two perspectives of proxy reporting of health-related quality of life using the Euroqol-5D, an investigation of agreement. Med Care. 2008;46:1140–8.

94. Ware JE, Kosinski M, Dewey JE, Gandek B. SF-36 health survey: manual and interpretation guide. Lincoln, RI: Quality Metric Inc.; 2002.

95. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993;85:365–76.

96. 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;13:171–87.

97. 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;20:1727–36.

98. Janssen M, Pickard AS, Golicki D, 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.

99. Malley J, Netten A. Measuring outcomes of social care. Res Policy Plan. 2009;27:85–96.

100. Simon J, Anand P, Gray A, Rugkåsa J, Yeeles K, Burns T. Operationalising the capability approach for outcome measurement in mental health research. Soc Sci Med. 2013;98:187–96.

101. Vergunst F, Jenkinson C, Burns T, Simon J. Application of Sen's capability approach to outcome measurement in mental health research: psychometric validation of a novel multi-dimensional instrument (OxCAP-MH). Hum Welfare. 2014;3:1–4.

102. Makai P, Brouwer WB, Koopmanschap MA, Stolk EA, Nieboer AP. 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.