

An Investigation of Construct Validity and Responsiveness of the Danish ICECAP-A

Annette Willemoes Holst-Kristensen (✉ awh@dcm.aau.dk)

Aalborg Universitet <https://orcid.org/0000-0002-6945-2553>

Paul Mark Mitchell

University of Bristol

Myles-Jay Linton

University of Bristol

Joanna Coast

University of Bristol

Kirsten Fonager

Aalborg Universitet

Kjeld Møller Pedersen

Aalborg Universitet

Research

Keywords: ICECAP-A, construct validity, responsiveness

Posted Date: June 11th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-32586/v1>

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

Abstract

Background

This study aimed to provide the first assessment of construct validity of ICECAP-A in patients with cardiovascular disease, chronic obstructive pulmonary disease and diabetes, and to assess the responsiveness of the measure in this group. Method

Data were provided from patients attending rehabilitation in the municipality of Aalborg, Denmark, from March 2018 to March 2019. Patients answered a questionnaire from the healthcare centre and the ICECAP-A at baseline and 12 weeks follow-up. To assess construct validity, *a priori* hypotheses were developed. Based on these hypotheses, associations between sociodemographic characteristics, 'general health', a freedom dimension, and ICECAP-A were analysed through chi-squared tests and Spearman rank correlations for categorical and ordinal variables, respectively. To investigate responsiveness, the anchor-based method was used. Patients were divided into improved, worsened or no change, based on changes between baseline and follow-up on the anchor measures ('general health' and 'freedom'). To quantify responsiveness, both the weighted and un-weighted ICECAP-A scores' effect sizes, standardised response means and t-tests were used. Findings were explored across different age groups.

Result

Of all the hypothesised associations, 16 of 26 (62%) were in the expected direction. The expected relationships were found between ICECAP-A scores and general health and freedom to do things. ICECAP-A was responsive in terms of capturing the effects on general health and the freedom to do things. Differences were found across age groups, with greater responsiveness to change in those aged under 65 years. The item-by-item analysis showed that capability was mainly driven by stability and autonomy.

Conclusion

This study has shown that the Danish ICECAP-A is a valid and responsive measure of the effects of an exercise and education-based rehabilitation programme.

Background

The ICECAP-A is a measure of wellbeing with a theoretical basis in Amartya Sen's work. The capability approach assesses wellbeing in terms of individual 'functionings' and 'capabilities'. Functionings refer to the things an individual 'is' or 'does', ranging from fundamental aspects of life such as 'being healthy' to more complex aspects such as 'having self-respect'. Capabilities represent an individual's freedom to carry out functionings. This is important, because a person may be able to function in a particular way, but may choose not to utilise that functioning [1–3].

The ICECAP-A conceptualises wellbeing as the capability of an individual to achieve valuable functionings. ICECAP-A has five attributes: stability, attachment, autonomy, achievement and enjoyment [4]. The initial aim of the ICECAP instruments was to develop a broad measure of quality of life (QoL) for use in economic evaluation [4]. Several other capability measures have been developed, such as the OxCAP and ASCOT [5–7]. However, the ICECAP measures are distinct as they provide a generic measure of capability wellbeing for use in the economic

evaluation of interventions in areas such as health and social care, where a broader aim like empowerment is to be explored [4, 8].

A systematic review by Afentou and Kinghorn [9] and a comparative review by Helter et al.[10] found evidence available on reliability, content validity, construct validity and responsiveness of the ICECAP-A measure in various populations, but so far, most evidence relates to the original UK version [9, 10]. In the Danish context, only one reliability study of the ICECAP-A in the general population exists [11]. The study reported here aims to provide the first assessment of construct validity in patients with cardiovascular disease (CVD), chronic obstructive pulmonary disease (COPD) and diabetes, and to assess the responsiveness of the ICECAP-A for this group in a Danish municipal rehabilitation setting.

Methods

Data collection and setting

Data were collected on a routine basis from patients attending rehabilitation in the municipality of Aalborg from March 2018 to April 2019. Patients were referred by their general practitioner or the hospital to the rehabilitation programme after an acute event necessitating a hospital stay related to their CVD, COPD or diabetes. Sociodemographic characteristics included age, gender (female or male), cohabitation (binary), education (defined in three levels according to the International Standard Classification of Education (ISCED): low < 11 years of schooling, medium 11–16 years of schooling, high > 16 years of schooling) and socioeconomic status (employed, unemployed or other benefits, or retired). All attending patients were asked to complete a questionnaire developed by the healthcare centre (the Aalborg questionnaire, available on request) and the ICECAP-A questionnaire at baseline and 12 weeks follow-up after the completion of the rehabilitation programme. It was the patient's choice as to whether they wished to complete the questionnaire on each occasion.

Municipal rehabilitation

In Denmark, the 98 municipalities offer rehabilitation programmes to chronically ill patients with, for example, CVD, COPD, and/or diabetes. The programmes are situated at the healthcare centre in Aalborg and at times in 'satellite' centres in varied locations across the municipality. The programmes provide exercise and education to groups of varying size. The exercise sessions take place one to two times a week and are of low to moderate intensity. The education component covers knowledge of the disease; dietary advice; the importance of physical activity, smoking cessation and medicine consumption; and goals and motivation. The programmes usually commence within a few weeks after discharge from the hospital and continue for 8–12 weeks [12].

Measuring rehabilitation outcomes

The municipality of Aalborg, Denmark, decided in 2018 to develop a self-completion questionnaire to evaluate their rehabilitation programme. The full questionnaire consists of 33 questions, including background information (gender, employment status, education level and cohabitation). Additional questions concerning training level and satisfaction with the program were asked at follow-up. The healthcare centre uses six of the questions to evaluate the rehabilitation programmes: (1) 'general health', (2) 'improvement of quality of life', (3) 'feeling fit to do the things I want to', (4) 'better at handling everyday life after programme', (5) 'know how to

sustain health in the future’ and (6) ‘able to be more physically active after programme’. Questions 1 and 3 were the only questions asked at both baseline and follow-up; the rest were only asked at follow-up. Questions 1–5 have four or five possible response categories (where higher scores indicate greater levels of general health, for example). Question 6 has a binary response option (yes or no).

Construct validity

Construct validity is the degree to which an instrument (such as a questionnaire) measures what it is hypothesised to be measuring. It can be assessed by considering the degree to which expected relationships between a measure and other factors are confirmed [13, 14]. Best-practice guidance on psychometric analyses highlights the importance of *a priori* statement of hypotheses on the anticipated relationship between the constructs explored [15]. Drawing on Sen’s theoretical framework for the establishment of capabilities, capability can be limited by reduced socioeconomic status and improved by good circumstances [3]. For the assessment of construct validity, *a priori* hypotheses were developed based on existing evidence about the ICECAP measures in other contexts [16, 17]. Table 1 indicates the expected direction between the five attributes of ICECAP-A, and indicators of socioeconomic status, general health and freedom in terms of ‘feeling fit to do the things I want to’ included in the Aalborg questionnaire.

Table 1
Hypothesised positive relationships between ICECAP-A attributes and the Aalborg questionnaire

ICECAP-A	Stability	Attachment	Autonomy	Achievement	Enjoyment	Total score
General health	+		+	+	+	+
+	+	+	+	+	+	
Employment	+	+	+	+	+	+
Education level	+		+	+		+
Cohabitation	+	+		+	+	+

The interpretation of Table 1 is as follows. The stability attribute is initially expressed as being able to feel settled and secure, and relates to the absence of significant changes in life and stress. It is therefore hypothesised that significant negative life changes were likely to be associated with reduced capability (such as changes in general health). The validity study by Al-Janabi et al. found that, among other factors, employment, education and relationship status were associated with stability in a positive direction [17]. Therefore, this study expected an association between stability and employment, education and cohabitation in a positive direction, despite the different definitions of relationship status and education level. The attachment attribute is stated in terms of being able to have love, friendship and support, and relates to the ability to interact with others and have good relationships. Al-Janabi et al. found an positive association between attachment, employment and relationship status [17]. This study therefore anticipated finding an association between attachment, employment and cohabitation in a positive direction. The autonomy attribute is defined as being able to be independent and relates to looking after oneself and making one’s own decisions. Previously, positive associations between autonomy and employment and education have been found [17]. It was therefore anticipated that higher capability level for autonomy would be associated with higher level of employment and education in this study. The achievement attribute is defined as being able to achieve and progress, and reflects individuals’ abilities to move forward and achieve their goals. Previously, positive associations between achievement and employment,

education and relationship status have been found [17]. It was therefore anticipated that capability for achievement would be associated with employment, education and cohabitation in a positive direction in this study. The enjoyment attribute is defined as being able to have enjoyment and pleasure in life. It reflects opportunities for the small pleasures in life, as well as things that are perceived to be enjoyable or exciting. As such, an association with employment and cohabitation was anticipated in a positive direction [17].

The ICECAP-A measure was developed to measure the effectiveness of health and social care interventions. The degree of variation in health and healthcare usage is reflected in individuals' capabilities, and therefore is essential and of interest, because poor health and disabilities affect one's capabilities [4, 17]. Previous studies concerning ICECAP-A have found that impairments to physical health reduce the capability for stability, autonomy, achievement and enjoyment [17, 18]. Therefore, this study anticipated an association between general health and stability, autonomy, achievement and enjoyment. Here, it was anticipated that the question focusing on general health would be interpreted by participants as a question about physical health only, given the reasons that they were accessing the service, and thus would not be associated with attachment. 'Feeling fit to do the things I want to' was hypothesised to be associated with all five attributes of the ICECAP-A, and high levels of capability were anticipated to relate to a high level of this question of freedom. This hypothesis is based on the findings by Al-Janabi et al. where a similar question was asked, 'I can do the things in life I want to do', and an association was found with all attributes [17].

Statistical analysis

Based on these hypotheses (Table 1), associations between selected variables and the ICECAP-A attributes at baseline were analysed using chi-squared tests for categorical variables and Spearman rank correlation for ordinal variables. A correlation was considered strong if the coefficient was higher than 0.5, moderate if the coefficient was between 0.3 and 0.5, and weak if the coefficient was below 0.3 [19].

Responsiveness

The ability of outcome measures to detect meaningful change, is central to their usefulness in health and social care interventions. Two core ideas in the assessment of evaluative instruments are sensitivity to change and responsiveness. Sensitivity to change refers to the ability of instruments to measure change statistically. Responsiveness addresses the detection of the clinically relevant change [13, 20].

To assess responsiveness, some criterion is needed to ascertain where patients have changed over time. The two main methods for assessing responsiveness are the distribution- and anchor-based approaches. The distribution-based method uses the effect size of the difference between groups to measure variability, standard response means, standard error of measurement and responsive statistics. The anchor-based method is sample-independent and examines the relationship with an anchor, such as a QoL measure, to explain the meaning of a particular degree of change [21]. The anchors can either be cross-sectional or longitudinal. An anchor-based analysis aims to assess whether scores on the target measure change in an anticipated way, as indicated by changes in the scores on the anchor [22]. Distribution methods alone do not provide information about the clinical relevance of the observed change. Therefore, this study assessed responsiveness, using anchor-based methods to investigate the association between change over time in the ICECAP-A scores and change over time

in the anchors. An exploratory analysis of the correlation between the change scores of longitudinal outcome measures was used to support the choice of anchors for this study.

Using Cohen's rule, correlations were considered strong when the coefficients were > 0.50 , moderate when ≥ 0.30 , and weak when < 0.30 . Therefore, 0.30 was used as a correlation threshold to define an at least moderate association between an anchor and outcome measure change score [23]. General health and 'feeling fit to do the things I want to' were the only two questions for which there were longitudinal data, but they were only used if they reached a threshold of baseline correlation of 0.3 (at least moderate correlation). For appropriate anchors, patients were divided into three groups depending on the changes in scores in general health and 'feeling fit to do the things I want to': (1) those who had worsened between baseline and follow-up scores, (2) those who had improved between baseline and follow-up scores, and (3) those with no change in scores between baseline and follow-up.

When assessing the responsiveness of a weighted measure such as ICECAP-A [8], consideration needs to be given independently to both the descriptive system [4] and the value weighting of the descriptive system. It is essential that the descriptive system can detect a change in a construct for the weighted measure to reflect meaningful change. If the analysis only uses the weighted tariff scores, a misleading conclusion could be made, that is, a conclusion whereby the measure is thought not to be responsive, when, in fact, the descriptive system of the measure shows change, but the value weightings suggest that these changes are not highly valued [24]. The weighted tariff scores are also reflective of the UK population and not those of the Danish public. Therefore, for each anchor, two analyses are presented: (1) an analysis of the 'un-weighted' descriptive system of the ICECAP-A and (2) an analysis of the 'weighted tariff scores'. For the un-weighted and weighted analysis, change was calculated in groups that improved and worsened. Un-weighted scores were calculated by summing ICECAP-A item response levels, with four indicating full capability on an item and one indicating no capability on an item. The weighted tariff scores were calculated using the UK general population tariff from Flynn et al. [25]. Findings were explored across different age groups (< 65 versus ≥ 65 years of age).

Responsiveness of the ICECAP-A scores was assessed using the Cohen's effect size (ES) and standardised response mean (SRM). Additionally, a paired t-test was applied to test the null hypothesis, that no change in the response means between baseline and follow-up had occurred. These indices were calculated separately for patients who reported improved, worsened or no change in the anchors [13, 23]. The effect size was calculated by dividing the mean difference between baseline and follow-up scores by the standard deviation (SD) of baseline scores; SRM was calculated by dividing the mean score change (follow-up minus baseline) by the standard deviation of the change [22]. For all indices, a value of < 0.2 was considered small, 0.2–0.5 moderate and > 0.5 large responsiveness [23]. The range of the un-weighted score was 16 (5–20), and for the tariff score was 1 (0–1) with higher scores on both representing higher capability. Age differences in responsiveness were investigated by subgroup analysis using a group < 65 years of age and a group ≥ 65 years of age.

To assess the responsiveness of the individual ICECAP-A items, a response profile (frequency of participants answering each level for each item, at baseline and follow-up) was completed for the two anchors. Change in response profiles between baseline and follow-up was analysed for each item to indicate which items were the 'drivers' of change in the overall measure.

Statistical analysis

The investigation of construct validity was based on all baseline data. The responsiveness analysis was based on complete cases in terms of questionnaire data because of high rates of missing data (78%); hence, imputation was not considered. The type of missing was anticipated to be missing completely at random because in all cases the entire questionnaire was missing. The reason for the amount of missing is that there was voluntary completion of the questionnaire, both at baseline and follow-up. Therefore, complete case analysis was performed for the responsiveness analysis. All analyses were carried out in Stata version 15 with a significance level set at 1% and 5%.

The study was carried out in accordance with the General Data Protection Regulation (2015-509-00007). In accordance with the Danish National Committee on Health Research Ethics, this research satisfies the criteria of being 'questionnaire and register-based research excluding human biological material', and thus was not required to undergo a formal ethics procedure [26].

Results

A total of 729 patients were registered at baseline as having completed the rehabilitation programme. At baseline, 454 patients completed the ICECAP-A. Of these, 155 completed the ICECAP-A at follow-up. The baseline characteristics for the complete cases and for the whole sample are presented in Table 2. More men were included, and just over half were aged over 65 years, with a similar proportion being retired. Around two thirds were living with a spouse and approximately half had a medium level of education, with a similar number having a low as a high level of education.

Table 2
Baseline characteristics

Characteristics	Category	Frequency (%) whole sample n = 729	Frequency (%) included in construct validity n = 454	Frequency (%) included in responsiveness n = 155
Gender				
	Female	305 (42%)	183 (40%)	61 (39%)
	Male	424 (58%)	271 (60%)	94 (61%)
Age				
	18–29	4(1%)	2(0.5%)	-
	30–44	31 (4%)	21(4.5%)	6 (4%)
	45–64	284 (39%)	186(41%)	72 (46%)
	65+	410 (56%)	245(54%)	77 (50%)
Occupation				
	Retired	407 (56%)	242(53%)	84 (54%)
	Employed	196 (27%)	142(31%)	51 (33%)
	Unemployed/ other benefits	126 (17%)	70(16%)	20 (13%)
Education (based on ISCED classification)				
	Low	193 (26%)	106(23%)	38 (24%)
	Medium	356 (49%)	222(49%)	63 (41%)
	High	180 (25%)	126(28%)	54 (35%)
Cohabiting				
	Cohabiting	483 (66%)	301(34%)	104 (67%)
	Non- cohabiting	246 (34%)	153(66%)	51 (33%)
Diagnosis				
	Diabetes	297 (41%)	166(36%)	56 (36%)
	Cardiovascular	215 (29%)	148(33%)	57 (37%)
	COPD	217 (30%)	140(31%)	42 (27%)
Baseline scores	Measure range			
Un-weighted score	5–20	16.63	16.63	16.65

Characteristics	Category	Frequency (%) whole sample n = 729	Frequency (%) included in construct validity n = 454	Frequency (%) included in responsiveness n = 155
Weighted tariff scores	0-1	0.87	0.87	0.88
General health	1-5	2.86	2.86	2.85
1-4	3.47	3.50	3.52	

Patients' responses (complete cases) at baseline and follow-up are listed in Table 3. The baseline weighted tariff score was 0.87 and the follow-up weighted tariff score was 0.89, thus a mean change of 0.02. The majority of responses had the highest or second-highest level of capabilities for each of the five attributes. Nevertheless, some patients indicated that their capability level was limited (little or no capability) in most of the five attributes. However, the proportion was small (< 5 patients), and in the autonomy attribute, there were no responses at the lowest level at follow-up. The percentage of patients reporting the highest response level increased for each of the attributes between baseline and follow-up data collection.

Table 3
Patient rehabilitation responses to ICECAP-A measure at baseline and 12 weeks follow-up.

ICECAP-A attributes (n = 155)	Baseline frequency (%)	Follow-up frequency (%)
Stability		
I am able to feel settled and secure in all areas of my life	56 (36%)	67 (43%)
I am able to feel settled and secure in many areas of my life	89 (57%)	79 (51%)
I am able to feel settled and secure in a few areas of my life	6 (4%)	8 (5%)
I am unable to feel settled and secure in any areas of my life	4 (3%)	1 (1%)
Attachment		
I can have a lot of love, friendship and support	82 (53%)	83 (54%)
I can have quite a lot of love, friendship and support	57 (36%)	58 (37%)
I can have a little love, friendship and support	15 (10%)	14 (9%)
I cannot have any love, friendship and support	1 (1%)	-
Autonomy		
I am able to be completely independent	77 (50%)	83 (54%)
I am able to be independent in many things	70 (45%)	67 (43%)
I am able to be independent in a few things	4 (2.5%)	5 (3%)
I am unable to be at all independent	4 (2.5%)	-
Achievement		
I can achieve and progress in all aspects of my life	30 (19%)	40 (26%)
I can achieve and progress in many aspects of my life	105 (68%)	106(68%)
I can achieve and progress in a few aspects of my life	16 (10%)	8 (5%)
I cannot achieve and progress in any aspects of my life	4 (3%)	1 (1%)
Enjoyment		
I can have a lot of enjoyment and pleasure	88 (57%)	99 (64%)
I can have quite a lot of enjoyment and pleasure	59 (38%)	49 (32%)
I can have a little enjoyment and pleasure	5 (3%)	7 (4%)
I cannot have any enjoyment and pleasure	3 (2%)	-

Construct validity

Table 4 shows the associations between selected variables and ICECAP-A attributes at baseline. Of the 26 hypothesised associations, 16 (62%) were in the expected direction. Hypothesised associations that did not meet our a priori tests were (1) education, cohabitation and the stability attribute, (2) employment and the attachment attribute, (3) employment, education (negative correlation, but close to zero – 0.0005) and the autonomy attribute, (4) education, cohabitation and the achievement attribute, (5) employment and the enjoyment attribute, and (6) employment, education and the weighted tariff score. In contrast, the associations between general health and the attachment attribute, were not hypothesised.

Table 4

Construct validity: association and correlation between ICECAP-A, baseline characteristics and questionnaire from the healthcare centre

	Stability	Attachment	Autonomy	Achievement	Enjoyment	Un-weighted score	Weighted tariff score
Association							
General health	0.00**	0.03*	0.00**	0.00**	0.00**	0.00**	0.00**
0.00**	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**	
Employment	0.00**	0.51	0.70	0.00**	0.13	0.26	0.15
Education	0.25	0.08	0.17	0.66	0.11	0.23	0.31
Cohabitation	0.09	0.00**	0.08	0.13	0.03*	0.00**	0.00**
Gender	0.78	0.28	0.14	0.25	0.35	0.06	0.07
Age	0.21	0.11	0.46	0.27	0.36	0.43	0.40
Correlation							
General health	0.51	0.33	0.31	0.43	0.46	0.54	0.50
0.50	0.28	0.34	0.39	0.44	0.52	0.52	

** Statistically significant (in expected direction) with p-values < 0.01

*Statistically significant (in expected direction) with p-values < 0.05

Bold = hypothesised

Responsiveness

Based on the correlations, analyses of general health and 'feeling fit to do the things I want to' were chosen as anchors, as both reached strong correlation (0.54 and 0.52) and were therefore appropriate to use as anchors (see Table 4). Table 5 shows the change in un-weighted and weighted tariff scores in groups that reported improved (n = 70) and worsened (n = 16) general health scores. In groups that reported improved general health scores, ICECAP-A scores increased (0.05), and in the groups that reported a worsening of general health scores,

ICECAP-A scores decreased (-0,06). The ES and SRM for those reporting an improvement in general health were small for both the un-weighted and weighted tariff scores; for those who reported a worsening in general health scores, the ES and SRM were moderate to strong. The ES and SRM in ICECAP-A scores were more substantial in the groups that reported a worsening of general health than improvement.

Table 5

Responsiveness: Mean changes in un-weighted scores and weighted tariff scores by anchor change groups (n = 155)

Anchor group	Baseline ICECAP-A (SD)	Follow-up ICECAP-A (SD)	Mean ICECAP-A change (95% CI)	Difference in SD	ES ^a	SRM ^b
General health						
Un-weighted scores						
Improved (n = 70)	16.39 (2.64)	17.53 (2.07)	1.14 (0.69;1.560)*	1.90	0.43	0.6
No change (n = 69)	16.76 (2.54)	16.88 (2.32)	0.12 (-0.22;0.46)	1.41	0.05	0.09
Worsened (n = 16)	17.25 (1.24)	16.19 (1.97)	-1.06 (-1.69;-0.43)*	1.18	-0.85	-0.90
Weighted tariff score						
Improved (n = 70)	0.87 (0.16)	0.92 (0.09)	0.05 (0.03;0.08)*	0.11	0.32	0.45
No change (n = 69)	0.88 (0.02)	0.89 (0.12)	0.01 (-0.01;0.03)	0.08	0.5	0.13
Worsened (n = 16)	0.92 (0.05)	0.86 (0.12)	-0.06 (-0.10;-0.01)*	0.09	-1.2	0.67
Un-weighted scores						
Improved (n = 37)	15.73 (2.75)	16.84 (2.29)	1.11 (0.54;1.67)*	1.70	0.40	0.65
No change (n = 103)	16.94 (2.45)	17.32 (2.17)	0.38 (0.04;0.72)*	1.74	0.16	0.22
Worsened (n = 15)	16.87 (1.50)	16.27 (2.09)	-0.6 (-1.51;0.31)	1.64	-0.40	0.37
Weighted tariff score						
Improved (n = 37)	0.83 (0.18)	0.89 (0.12)	0.06 (0.03;0.10)*	0.10	0.33	0.6
No change (n = 103)	0.89 (0.14)	0.91 (0.11)	0.02 (-0.001;0.04)	0.09	0.14	0.22
Worsened (n = 15)	0.90 (0.05)	0.87 (0.10)	-0.03 (-0.08;0.01)	0.08	-0.60	0.38

*Statistically significant with p-values < 0.05

^aES (effect size) – mean ICECAP-A change/SD of baseline scores

^bSRM (standardised response mean) – mean change/SD of the difference

Table 5 shows the change in un-weighted and weighted tariff scores in groups that reported improved (n = 37) and worsened (n = 15) 'freedom' scores. In groups that reported improved freedom scores, ICECAP-A scores increased (0.06), and in the groups that reported a worsening of freedom scores, ICECAP-A scores decreased (-0.03). The change in ICECAP-A scores was more substantial in the groups that reported an improvement of freedom. The ES and SRM for those reporting an improvement in freedom were small to moderate for both the un-weighted and weighted tariff scores; for those who reported a worsening in freedom scores, the ES and SRM were small.

Subgroup analysis of responsiveness in different age groups

The results concerning responsiveness in the different age groups (Table 6) showed small differences, with the younger age group having a higher mean change, ES and SRM than the older group. In anchor group GH < 65 the improved patients had a weighted tariff score of 0.85 at baseline and 0.91 at follow-up – mean change 0.06. The worsened group < 65, had a weighted tariff score of 0.91 at baseline and 0.85 at follow-up – mean change – 0.06. In the group > 65 the patients had a weighted tariff score of 0.88 at baseline and 0.93 at follow-up – mean change 0.05 in the improved group. The worsened group ≥ 65, had a weighted tariff score of 0.91 at baseline and 0.86 at follow-up – mean change – 0.05. In the anchor group 'Feeling fit to do the things I want to' the patients that improved had a weighted tariff score of 0.82 at baseline and 0.88 at follow-up – mean change 0.06. The worsened group had a weighted tariff score of 0.89 at baseline and 0.82 at follow-up – mean change – 0.07. In the group ≥ 65 with GH as anchor the patients that improved had a weighted tariff score of 0.83 at baseline and 0.91 at follow-up – mean change 0.08. The worsened group ≥ 65, had a weighted tariff score of 0.92 at baseline and 0.93 at follow-up – mean change – 0.0003. The sample size was small; however, the distribution was 50/50 between groups. More respondents reported improved general health (n = 35) compared with those improving in 'Feeling fit to do the things I want to' (n = 15). The ES and SRM were larger in the < 65 groups. In the < 65 group, both the improved and worsened mean change were statistically significantly different between baseline and follow-up. This was only the case with the improved group in the ≥ 65 subgroup. Results concerning freedom showed small ES and SRM in both age groups, but smallest in the ≥ 65 subgroup.

Table 6

Responsiveness: Mean changes in un-weighted scores and weighted tariff scores by age groups

Anchor group	Baseline ICECAP-A (SD)	Follow-up ICECAP-A (SD)	Mean ICECAP-A change (95% CI)	Difference in SD	ES ^a	SRM ^b
General health						
Age group under 65						
Un-weighted scores						
Improved (n = 35)	16.03 (2.97)	17.46 (2.42)	1.43* (0.80;2.06)	1.85	0.48	0.77
No change n = 33)	16.57 (2.62)	16.60 (2.52)	0.03 (-0.39;0.44)	1.16	0.01	0.02
Worsened n = 10)	17.3 (1.34)	16.1 (1.91)	-1.2* (-2.08;-0.32)	1.23	-0.90	-0.98
Weighted tariff score						
Improved (n = 35)	0.85 (0.19)	0.91 (0.12)	0.06* (0.02;0.10)	0.11	0.32	0.55
No change (n = 33)	0.86 (0.15)	0.87 (0.14)	0.01 (-0.01;0.03)	0.55	0.07	0.02
Worsened (n = 10)	0.91 (0.05)	0.85 (0.11)	-0.06* (-0.12;-0.004)	0.08	-1.2	0.75
Age group 65+						
Un-weighted scores						
Improved (n = 35)	16.74 (2.27)	17.6 (1.68)	0.85* (0.20;1.52)	1.93	0.37	0.44
No change (n = 36)	16.94 (2.48)	17.14 (2.11)	0.19 (-0.35;0.74)	1.62	0.08	0.12
Worsened (n = 6)	17.17 (1.17)	16.33 (2.25)	-0.83 (-2.06;0.40)	1.60	-0.71	-0.52
Weighted tariff score						
Improved (n = 35)	0.88 (0.12)	0.93 (0.05)	0.05* (0.01;0.08)	0.10	0.42	0.5
No change (n = 36)	0.89 (0.13)	0.90 (0.11)	0.01 (-0.02;0.04)	0.11	0.08	0.1
Worsened (n = 6)	0.91 (0.04)	0.86 (0.14)	-0.05 (-0.16;0.06)	0.09	-1.3	-0.6

Anchor group	Baseline ICECAP-A (SD)	Follow-up ICECAP-A (SD)	Mean ICECAP-A change (95% CI)	Difference in SD	ES ^a	SRM ^b
Age group under 65						
Un-weighted scores						
Improved (n = 22)	15.55 (2.84)	16.68 (2.71)	1.14* (0.55;1.72)	1.32	0.40	0.86
No change (n = 48)	16.81 (2.65)	17.31 (2.25)	0.50 (-0.02;1.02)	1.80	0.19	0.28
Worsened (n = 8)	16.5 (1.93)	15.25 (2.12)	-1.25 (-2.57;0.07)	1.58	-0.65	-0.79
Weighted tariff score						
Improved (n = 22)	0.82 (0.19)	0.88 (0.14)	0.06* (0.02;0.08)	0.07	0.32	0.71
No change (n = 48)	0.88 (0.16)	0.90 (0.12)	0.02 (-0.00;0.05)	0.1	0.13	0.1
Worsened (n = 8)	0.89 (0.07)	0.82 (0.12)	-0.07 (-0.14;0.00)	0.08	-1	-0.38
Age group 65+						
Un-weighted scores						
Improved (n = 15)	16(2.67)	17.06 (1.53)	1.06 (-0.14;2.28)	2.19	0.40	0.48
No change (n = 55)	17.05(2.12)	17.33(2.12)	0.27 (-0.19;0.73)	1.69	0.13	0.39
Worsened (n = 7)	17.29 (0.76)	17.43 (1.40)	0.14 (-1.21;1.50)	1.46	0.18	0.1
Weighted tariff score						
Improved (n = 15)	0.83 (0.16)	0.91 (0.06)	0.08 (-0.00;0.16)	0.14	0.50	0.57
No change (n = 55)	0.90 (0.11)	0.91 (0.10)	0.01 (-0.01;0.03)	0.09	0.50	0.11
Worsened (n = 7)	0.92 (0.03)	0.93 (0.04)	0.003 (-0.03;0.04)	0.04	0.10	0.08

*Statistically significant with p-values < 0.05

^aES (Effect size) – mean ICECAP-A change/SD of baseline scores

^bSRM (Standardised response mean) – mean change/SD of the difference

The item-by-item analysis (Table 7) showed that in the group of patients reporting an improvement in general health, the largest increase was in stability and in the patients reporting worsening of general health, the biggest decrease was in autonomy. In the group of patients reporting an improvement in ‘feeling fit to do the things I want to’, the increase was comparable across attributes with increases in attachment lowest, and in the patients reporting worsening in ‘feeling fit to do the things I want to’, the biggest decreases were seen in autonomy.

Table 7 Item-by-item analysis: Distribution of changed response according to anchor

Anchor		Change between baseline and follow-up (%)				
		Stability	Attachment	Autonomy	Achievement	Enjoyment
General health						
	Improved (n=70)	↑ 15 (21) ↔ 55 (79) ↓ -	3 (4) 67 (96) -	9 (13) 61 (87) -	12 (17) 68 (83) -	10 (14) 60 (86) -
	Worsened (n=16)	↑ - ↔ 13 (81) ↓ 3 (19)	2 (12.5) 12 (75) 2 (12.5)	- 11 (69) 5 (31)	- 14 (88) 2 (12)	- 13 (81) 3 (19)
‘Feeling fit to do the things I want to’						
	Improved (n=37)	↑ 5 (14) ↔ 32 (86) ↓ -	3 (8) 34 (92) -	6 (16) 31 (84) -	6 (16) 31 (84) -	5 (14) 32 (86) -
	Worsened (n=15)	↑ - ↔ 14 (93) ↓ 1 (7)	- 13 (87) 2 (13)	- 10 (67) 5 (33)	- 14 (93) 1 (7)	- 15 (100) -

Discussion

This is the first study to assess the construct validity and responsiveness of the Danish ICECAP-A measure. To achieve this, it used longitudinal data from a rehabilitation setting in a population of chronically ill patients. The findings indicate that scores on the Danish ICECAP-A are associated with indicators of freedom and general health. The results provide evidence about the instrument’s ability to respond to differences in socioeconomic characteristics such as employment, education and cohabitation. The responsiveness analysis explored changes in the ICECAP-A scores in response to general health and freedom, and the results indicate that the ICECAP-A is responsive and that patients younger than 65 years of age appear more responsive than older patients. The Danish ICECAP-A, therefore, demonstrated encouraging construct validity and responsiveness in a rehabilitation setting among chronically ill patients. The item-by-item analysis showed that those reporting an increase in general health and ‘Feeling fit to do the things I want to’ scores the largest change in Achievement and autonomy respectively, and those reporting an decreased general health and ‘Feeling fit to do the things I want to’ score the largest change was found in autonomy in both.

The overall findings are consistent with previous studies that found the ICECAP-A to be promising in terms of validity [16, 17, 27] and responsiveness [16, 28] in different populations and health conditions. The most comparable is the study by Al-Janabi et al. [17], where the ICECAP-A was found to be associated with various socioeconomic variables, the EQ-5D, and questions concerning freedom and opportunities. The most noticeable result was that the present study found an association between general health and the attribute attachment where Al-Janabi et al. found the opposite. Al-Janabi et al. did, however, find an association between anxiety and

depression and attachment. This could indicate that the participants in this study considered mental health to be a part of general health, which could relate to differences in the setting, but could also reflect the increasing focus on mental health across society more generally since the Al-Janabi research was published in 2013.

The study benefits from the available Danish ICECAP-A translation (discussed elsewhere [11]) that made it possible to investigate the psychometric properties of ICECAP-A. Further, this study extends our academic knowledge around accurate outcomes assessment in the context of rehabilitation medicine among chronically ill patients. ICECAP-A is still a relatively new questionnaire, and so developing a better understanding of the tool's validity and responsiveness across populations is essential for its further use in health economic evaluations. Previous studies have demonstrated construct validity in different populations, including the general British population [17], women with irritable lower urinary tract symptoms [16] and a population with depression [27].

One methodological limitation of the study is the small number of possible anchors and lack of clinical anchors. While the use of general health as an anchor was driven by methodological considerations when considering a capability measure's suitability for use in health interventions, it is essential to identify how the instrument responds to changes in health. Health is one of many factors that affect the capability of a person and a relevant factor in this study population in particular. A smaller change in capability scores would, therefore, be expected in response to changes in health, and could have been useful to investigate with more anchors than general health. A previous study used EQ-5D as an anchor, in a population with depression, resulting in a correlation between all attributes of the ICECAP-A [27]. This study had a large proportion of missing data in terms of patients not having both a baseline and follow-up measures. This missingness was anticipated to be completely at random because the entire questionnaires were missing, rather than responses to specific questions. The amount of missingness may be due to that fact that it was voluntary completion of the questionnaire, both at baseline and follow-up. This could influence the results if the sample is different from the missing data and decrease the power of the sample. However, the proportion of missing was assumed too large (78%) to impute.

The evidence of validity and responsiveness presented in this study adds to the psychometric profile of the ICECAP-A measure, and the results provide an initial indication that the ICECAP-A may be responsive in public health research and chronically ill populations. In the Danish municipal rehabilitation setting, no national outcome measurement procedures exist, so a more extensive study with more participating municipalities would be interesting to explore the implications further. Establishing the psychometric performance of a measure is a continuous process, and further research is needed to explore how well the ICECAP-A performs in different public health and social care settings, such as in interventions regarding self-care. Ideally, capability measures could be incorporated into future health agreements and clinical guidelines. More importantly, it is necessary to show personnel in healthcare centres and decision-makers the benefits of implementing ICECAP-A in everyday work as a tool in public health and social care interventions, and not just as a scientific instrument.

Conclusion

This study provides the first investigation into construct validity and responsiveness to change for the Danish translation of the ICECAP-A and the first investigation into responsiveness to change for any ICECAP measure in the context of CVD, COPD and diabetes. The Danish ICECAP-A has demonstrable potential for accurately measuring the effect of rehabilitation. Furthermore, it appears to be responsive in terms of capturing the effects on general health and the freedom to do things. Future research into the psychometric properties of the Danish

ICECAP-A would be beneficial to clinicians and decision-makers in Denmark interested in capturing broader benefits to patients, beyond just health.

Declarations

Ethics approval and consent to participate

The study was carried out in accordance with the General Data Protection Regulation (2015-509-00007). In accordance with the Danish National Committee on Health Research Ethics, this research satisfies the criteria of being 'questionnaire and register-based research excluding human biological material', and thus was not required to undergo a formal ethics procedure [29].

Consent for publication

Not applicable

Availability of data and materials:

The data that support the findings of this study are available from [Denmark Statistics and the municipality of Aalborg, Denmark] but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of [Denmark Statistics and the municipality of Aalborg, Denmark].

Competing interests: Joanna Coast led the development of the ICECAP-A. The remaining authors declare that they have no conflict of interest.

Funding: This study was funded by the municipality of Aalborg, and The Health Foundation (Helse Fonden) (grant number 463019). Joanna Coast and Paul Mitchell's contribution to this work was funded by Wellcome [205384/Z/16/Z].

Authors' contributions: AWHK, PMM, MJL, JC, KF and KMP contributed to the conception and design of the study, MJL contributed with expertise to methodological discussions. AWHK conducted the analyses and drafted the manuscript. AWHK, PMM, MJL, JC, KF and KMP contributed to the interpretation of the work. AWHK, PMM, MJL, JC, KF and KMP critically revised the manuscript. AWHK, PMM, MJL, JC, KF and KMP gave final approval of the manuscript and agreed to be accountable for all aspects of the work ensuring integrity and accuracy.

Acknowledgements: We would like to thank all those who took part in the research, the healthcare centre in Aalborg and the participants in the rehabilitation programme.

References

1. Sen A. *Inequality Reexamined*. Harvard University Press, 1997.
2. Robeyns I. *Wellbeing, Freedom and Social Justice - The Capability Approach Re-Examined*. Cambridge: Open Book Publisher; 2017.
3. Sen A. Capability and well-being. In: M. Nussbaum & A. Sen, editors, *Quality of Life*. Oxford: Oxford University Press; 1993. pp. 30–53.

4. Al-Janabi H, Flynn TN, Coast J. Development of a self-report measure of capability wellbeing for adults: The ICECAP-A. *Qual Life Res.* 2012;21:167–76.
5. Lorgelly PK, Lorimer K, Fenwick EAL, et al. Operationalising the capability approach as an outcome measure in public health: The development of the OCAP-18. *Soc Sci Med.* 2015;142:68–81.
6. Simon J, Anand P, Gray A, et al. Operationalising the capability approach for outcome measurement in mental health research. *Soc Sci Med.* 2013;98:187–96.
7. Netten A, Burge P, Malley J, et al. Outcomes of social care for adults: Developing a preference-weighted measure. *Health Technol Assess (Rockv).* 2012;16:1–165.
8. 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.
9. Afentou N, Kinghorn P. Systematic Literature Review A Systematic Review of the Feasibility and Psychometric Properties of the ICEpop CAPability Measure for Adults and Its Use So Far in Economic Evaluation. *Value Heal.* 2020;23:515–26.
10. Mariann Helter T, Coast J, Łaszewska A, et al. Capability instruments in economic evaluations of health-related interventions: a comparative review of the literature. *Qual Life Res*; 1: 3.
11. Holst-Kristensen AW, Fonager K, Pedersen KM. Test–retest reliability of ICECAP–A in the adult Danish population. *Qual Life Res* 2019; 547–557.
12. Johansen J, Rahbek J, Møller K, et al. *Hvidbog om rehabiliteringsbegrebet. Rehabilitering i Danmark*, http://www.marselisborgcentret.dk/fileadmin/filer/Publikationer/PDF_er/Hvidbog.pdf (2004).
13. Streiner DL, Norman GR, Cairney J. *Health measurement scales, a practical guide to their development and use*. 5th ed. Oxford University Press, 2015.
14. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull.* 1955;52:281–302.
15. Kane M. Current Concerns in Validity Theory. *J Educ Meas.* 2001;38:319–42.
16. Goranitis I, Coast J, Al-Janabi H, et al. The validity and responsiveness of the ICECAP-A capability-well-being measure in women with irritative lower urinary tract symptoms. *Qual Life Res.* 2016;25:2063–75.
17. Al-Janabi H, Peters TJ, Brazier J, et al. An investigation of the construct validity of the ICECAP-A capability measure. *Qual Life Res.* 2012;22:1831–40.
18. Flynn TN, Chan P, Coast J, et al. Assessing Quality of Life among British Older People Using the ICEPOP CAPability (ICECAP-O) Measure. 2011; 9: 317–329.
19. Cohen J. Set Correlation and Contingency Tables. *Appl Psychol Meas.* 1988;12:425–34.
20. Brazier J, Ratcliffe J, Salomon JA, et al. *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford University Press, 2007.
21. Jayadevappa R, Malkowicz SB, Wittink M, et al. Comparison of distribution- and anchor-based approaches to infer changes in health-related quality of life of prostate cancer survivors. *Health Serv Res.* 2012;47:1902–25.
22. Wyrwich KW, Norquist JM, Lenderking WR, et al. Methods for interpreting change over time in patient-reported outcome measures. *Qual Life Res.* 2013;22:475–83.
23. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Lawrence Erlbaum Associates, 1988.

24. Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. *Health Econ.* 1999;8:41–51.
25. Flynn TN, Huynh E, Peters TJ, et al. Scoring the Iccap-a capability instrument. Estimation of a UK general population tariff. *Health Econ.* 2015;24:258–69.
26. Ministry of Health. *Lov om videnskabsetisk behandling af sundhedsvidenskabelige forskningsprojekter*, <https://www.retsinformation.dk/pdfPrint.aspx?id=192671> (2017, accessed 28 October 2019).
27. Mitchell PM, Al-Janabi H, Byford S, et al. Assessing the validity of the ICECAP-A capability measure for adults with depression. *BMC Psychiatry*, 17:46.
28. Keeley T, Al-Janabi H, Nicholls E, et al. A longitudinal assessment of the responsiveness of the ICECAP-A in a randomised controlled trial of a knee pain intervention. *Qual Life Res.* 2015;24:2319–31.