

Measuring health and broader wellbeing benefits in the context of opiate dependence: the psychometric performance of the ICECAP-A and EQ-5D-5L

Running title: Using ICECAP-A and EQ-5D-5L in drug addiction

Ilias Goranitis, PhD¹, Joanna Coast, PhD², Ed Day, PhD^{3,4}, Alex Copello, PhD^{3,5}, Nick Freemantle, PhD⁶, Jennifer Seddon, PhD³, Carmel Bennett, PhD³, Emma Frew, PhD^{1,*}

¹Health Economics Unit, Institute of Applied Health Research, University of Birmingham, UK; ²School of Social and Community Medicine, University of Bristol, UK; ³Birmingham & Solihull Mental Health NHS Foundation Trust, Research & Innovation Department, UK; ⁴Addictions Department, Institute of Psychiatry, Psychology & Neuroscience, King's College London, UK; ⁵School of Psychology, University of Birmingham, UK; ⁶Department of Primary Care and Population Health, University College London, UK

*Correspondence to: Emma Frew, Health Economics Unit, Public Health building, University of Birmingham, Birmingham, B15 2TT, UK. E-mail address: E.Frew@bham.ac.uk

Funding: The study was funded by the National Institute for Health Research (NIHR) under its Research for Patient Benefit (RfPB) Programme (Grant Reference Number PB-PG-0610-22392).

Acknowledgements: Early draft of this paper was presented at the Health Economists' Study Group (HESG) meeting hosted by the Health Economics at Lancaster Research Group at Lancaster University. We would like to acknowledge the feedback and the useful suggestions from those who participated in the session, Dr. Louise Longworth, Dr. Paula Lorgelly, Dr. Claudia Pereira, and particularly the paper discussant, Dr. James Shearer. We are also grateful to all the participants who took part in the study and to the community drug and alcohol teams located at Leicester (Leicestershire Partnership NHS Trust) and Solihull (Birmingham & Solihull Mental Health NHS Foundation Trust). This work represents independent research funded by the National Institute for Health Research (NIHR), and the views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Conflict of interest: J.C. was involved in the development of the ICECAP-A.

Trial registration: ISRCTN22608399

Keywords: ICECAP, EQ-5D, construct validity, sensitivity to change, addiction, mental health, outcome valuation, economic evaluation

Key points

What is already known about the topic?

- Treatments for substance use disorders aim at improving health and, most importantly, broader wellbeing.
- Health-related measures of quality of life may not adequately capture the broader wellbeing benefits associated with such complex interventions.
- Comprehensive evidence on the validity and responsiveness of these measures in this context is lacking.
- When broader capability-wellbeing benefits are expected, NICE recommends the parallel use of an ICECAP measure but their measurement properties in this context are yet unknown.

What does the paper add to existing knowledge?

- This is the first study exploring the psychometric properties of the EQ-5D-5L and ICECAP-A measures in the context of drug addiction.
- The results indicate that the ICECAP-A is a promising measure for capturing both health and broader wellbeing benefits in this context.

ABSTRACT

Background: Measuring outcomes in economic evaluations of social care interventions is challenging, as both health and wellbeing benefits are evident. The ICEpop CAPability instrument for Adults (ICECAP-A) and the five-level EuroQol-5D (EQ-5D-5L) are measures potentially suitable for the economic evaluation of treatments for substance use disorders. However, evidence for their validity in this context is lacking. **Objective:** To assess the construct validity of the ICECAP-A and EQ-5D-5L, in terms of convergent and discriminative validity, and sensitivity to change based on standard clinical measures (CORE-OM, TOP, ISEL, LDQ, and SSQ). **Methods:** Secondary analysis of pilot trial data for heroin users in opiate substitution treatment. Baseline convergence with clinical measures was assessed using Pearson's correlation coefficient. Discriminative validity was assessed using one-way ANOVA and stepwise regressions. Sensitivity to changes in clinical indicators were assessed at 3 and 12 months using the standardised response mean statistic, and parametric and non-parametric testing. **Results:** Both measures had the same level of construct validity, except for clinical indicators of wellbeing, where the ICECAP-A performed better. The ICECAP-A was sensitive to changes in both health and wellbeing indicators. The EQ-5D-5L had lower levels of sensitivity to change, and a ceiling effect (27%), particularly evident in the dimensions of *self-care* (89%), *mobility* (75%), and *usual activities* (72%). **Conclusions:** The findings support the construct validity of both measures, but ICECAP-A gives more attention to broader impacts, and is more sensitive to change. The ICECAP-A shows promise in evaluating treatments for substance use disorders where recovery is the desired outcome.

Introduction

Generic preference-based measures of health-related quality of life (HrQoL) are well-established in the economic evaluation of health care interventions. Policy makers, such as the National Institute for Health and Care Excellence (NICE) in the UK [1], have advocated their use in health technology assessment (HTA) as they allow for the direct comparison of cost-effectiveness across interventions [2]. Within the UK, the EuroQol-5D (EQ-5D) is the most commonly used preference-based measure [3], and the one recommended by NICE for HTA submissions [4]. Since 2011, a five-level version of the EQ-5D (EQ-5D-5L) has replaced the previous three-level version (EQ-5D-3L) [5], and emerging evidence indicates improved psychometric performance [6-9].

The EQ-5D-5L describes HrQoL through dimensions of *mobility, self-care, usual activities, pain/discomfort, and anxiety/depression*. It is designed to measure a core set of domains believed to impact on quality of life and linked to health. However, with this focus on certain aspects of health, the EQ-5D may not adequately reflect the impact of all health care interventions [10], such as for hearing or vision disorders [11], and severe mental health problems [12], and particularly of those resulting in broader personal and interpersonal wellbeing benefits, such as public health and social care interventions [13-19]. An alternative framework for measuring treatment benefits is with the ICEpop CAPability (ICECAP) measures, underpinned by Amartya Sen's capability approach [20]. The ICECAP instruments are designed to measure a particular set of capabilities with less emphasis on *actual* functioning, and more emphasis on the *ability* to function. These measures are focused on a wider notion of wellbeing and can be used in economic evaluations. To date, the ICECAP-O (older population) [21], ICECAP-A (adult population) [22], and ICECAP-SCM (supportive care) [23] have been developed.

Opiates carry the highest burden of disease of any drug of dependence, and are a major cause of public health problems [24]. In an era of austerity coupled with increasing health care costs, it is important that cost-utility analyses are available to guide any evaluation of treatments for opiate use

disorders. One of the challenges of cost-effectiveness analyses in services for substance use disorders is that they are intended primarily for single outcome treatments, but substance use disorder treatment results in a variety of outcomes [25]. Purchasers of treatment services are increasingly asking “*to what extent is your service making peoples’ lives better, rather than simply suppressing their drug use?*” [26, 27], and the process of ‘*Recovery*’ incorporates not just control over drug use, but also physical and mental health, and participating in meaningful roles within society [28-30].

In such a context, there is need to capture broader aspects of wellbeing than health status [26], and NICE recommends a parallel use of the EQ-5D and ICECAP measures [31]. Despite the positive emerging evidence for the psychometric properties of the ICECAP-A [32-35] and EQ-5D-5L [6-9], no evidence is yet available for the context of opiate dependence. Even though some supporting evidence for the EQ-5D-3L and other similar health-focused measures exists for this context [36-39], such measures are used with caution in economic evaluations [40, 41]. This study aims to explore the psychometric performance of the ICECAP-A and EQ-5D-5L in a sample of heroin users in opiate substitution treatment.

Methods

Data source

The study relies on a secondary analysis of data drawn from a pilot randomised controlled trial designed to investigate the clinical and cost-effectiveness of two psychological interventions delivered as an adjunct to usual care of individuals who had been receiving opiate substitution treatment for more than 12 months. The full trial protocol can be found elsewhere [42]. The two psychological interventions were a brief adapted version of ‘*Social Behaviour and Network Therapy (B-SBNT)*’, which actively aimed to involve social network members in the process of helping participants achieve their personal goals [43, 44], and ‘*Personal Goal Setting*’, which set specific goals and monitored achievement on an individual basis [45]. Patients were recruited if they were receiving

opiate substitution treatment with methadone or buprenorphine for at least a year but still reported heroin use within the past month. Individuals with concurrent severe mental or physical illness, depression, or legal problems were excluded. Interventions were delivered by three NHS specialist drug treatment services in England and participants were followed-up for 1 year. Participants provided written informed consent and ethical approval was granted by the Black Country NHS Research Ethics Committee (Reference: 12/WM/0046).

Outcome measures

Participants were assessed in terms of substance use (TOP) [46], mental health (CORE-OM) [47], social support (ISEL) [48], substance dependence (LDQ) [49], social satisfaction (SSQ) [50], capability-wellbeing (ICECAP-A) [22], and HrQoL (EQ-5D-5L) [5]. Questionnaires were completed during face-to-face interviews at baseline, and at 3 months and 12 months post-randomisation. Further information about each measure is presented below.

Treatment Outcomes Profile (TOP)

The TOP is used for monitoring substance use treatment and records information about substance use, injecting risk behaviour, offending, and health and social functioning over the past four weeks [46]. The latter category includes participants' subjective rating of psychological and physical health status, and overall quality of life on a 0 (poor) to 20 (good) scale. These dimensions were all included in the psychometric analysis.

Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM)

The CORE-OM is used to assess mental health problem severity and evaluate the impact of psychological interventions [47, 51]. It comprises 34 items grouped into the dimensions of wellbeing, symptoms (i.e. anxiety, depression, physical symptoms, and trauma), functioning (i.e. closeness,

general and social functioning), and risk (i.e. self-risk and risk to others). Each item is scored on a 0-4 scale, whereby a higher score denotes increased severity.

Interpersonal Support Evaluation List (ISEL)

The ISEL is used to measure perceived availability of social support in terms of material aid (tangible support), presence of people to discuss personal issues (appraisal support), presence of people to be favourably compared with (self-esteem support), and presence of people to socialize with (belonging support) [48, 52]. Each dimension of support comprises 10 items scored on a 0-3 scale, with higher values indicating greater support. Dimension scores are calculated by summing item responses and range from 0 to 30.

Leeds Dependence Questionnaire (LDQ)

The LDQ is used to measure dependence on a range of substances [49]. It comprises 10 items and each item is scored on a 0-3 scale, with higher scores indicating higher levels of substance dependence. A total score, ranging from 0 to 30, is calculated by summing all items, with the values of 10 and 20 representing the cut-off points for moderate and high level of dependence.

Social Satisfaction Questionnaire (SSQ)

The SSQ is used to assess satisfaction with the social environment in people with substance use disorders [50, 53]. It comprises 8 items scored on a 1-4 scale, with higher scores denoting higher level of social satisfaction. An overall index score is calculated by summing the score of all items and ranges from 8 to 32.

ICEpop CAPability measure for Adults (ICECAP-A)

The ICECAP-A is a measure of capability-wellbeing and captures the capability of individuals to function in terms of *stability, attachment, autonomy, achievement, and enjoyment* [22]. Each attribute

is scored on a 1-4 scale, with higher values indicating higher capability. A capability index score is assigned to the 1,024 (4^5) different response permutations using a scoring algorithm based on preferences elicited from a sample of 413 members of the UK general population with the best-worst scaling method [54]. Capability index scores range from 0 to 1, depending on whether no capability or full capability is reported across all five attributes of the ICECAP-A.

EuroQol five dimensional and five level questionnaire (EQ-5D-5L)

The EQ-5D is a measure of HrQoL and captures health status in terms of *mobility, self-care, usual activities, pain/discomfort, and anxiety/depression* [3]. The EQ-5D-5L dimensions are scored on a 1-5 scale, with higher values indicating extreme health problems [5]. The different combinations of individual responses are used to generate 3,125 (5^5) health states. A health index score is assigned to each health state using country-specific value sets. An EQ-5D-5L value set for England has recently become available [55]. This value set, herein referred to as “EQ-5D-5L England”, was developed based on the time trade-off (TTO) and discrete choice experiment (DCE) methods from a sample of 996 adult members of the general population. Health index scores range from -0.281 to 1, with negative values representing health states worse than death, 0 representing the ‘dead’ state, and 1 the ‘full health’ state. Until the publication of this value set, studies published in the UK, and elsewhere, on the EQ-5D-5L were relying on a crosswalk health index score calculated using a mapping algorithm to the UK general population tariff developed for the EQ-5D-3L based on the TTO method alone from a sample of 3,395 members [56, 57]. In this value set, herein referred to as “EQ-5D-3L UK crosswalk”, health index scores range from -0.594 to 1, with 26.7% of health states being considered worse than death. In the EQ-5D-5L England, these health states correspond to only 4.9% of all health states [55].

Analyses

The analysis involved an assessment of construct validity and sensitivity to change of the ICECAP-A and EQ-5D-5L. Analyses using the EQ-5D-5L relied on the EQ-5D-5L England value set. For

completeness, a supplementary analysis was undertaken using the EQ-5D-3LUK crosswalk value set. Given that the aim of the study was to explore actual completions of the questionnaires in order to ascertain whether the questionnaires work, no data imputation was performed, and analyses relied on available cases.

Construct validity is the ability of an instrument to measure the construct that it was designed to measure. Two forms of construct validity are convergent and discriminative validity [58].

Convergent validity assesses whether a measure and other variables or measures of the same or an overlapping construct are related to each other as expected [58]. The convergence of the ICECAP-A and EQ-5D-5L index scores with the five clinical measures was assessed based on baseline data using the Pearson correlation coefficient (r). Correlations were considered strong if $r > 0.5$, moderate if $0.3 \leq r \leq 0.5$, and weak if $r < 0.3$ [59].

Discriminative validity assesses the extent to which a measure is able to distinguish between dissimilar constructs [58]. A range of constructs were developed based on the different dimensions of the five clinical measures. Where clinical cut-off points were available, these were used to develop constructs. Alternatively, constructs were developed by splitting the sample into those above and below the mean score of the relevant clinical measure. A univariate analysis using one-way ANOVA was initially undertaken to explore discriminative validity based on baseline data. To avoid potential confounding problems associated with univariate analyses, a multivariate analysis was additionally carried out using stepwise multiple linear regressions. Age, gender, marital status, ethnicity, educational and employment status, type of accommodation, receiving state benefits, length of opiate substitution treatment, length of heroin abstinence, and the intervention-arm were used as model covariates if they recorded a p -value of ≤ 0.3 .

Due to the small number of participants achieving clinically important changes in some measures [60], and the lack of clinically important thresholds of change for all measures, the analysis focused on the assessment of sensitivity to change. In the assessment of sensitivity to change, the different

measures are commonly compared for patient groups expected to have experienced a change in health and wellbeing based on an external criterion (anchor) [61]. In this analysis, changes in the ICECAP-A and EQ-5D-5L index scores were evaluated on the basis of overall changes in the number of days of abstinence from heroin and changes in the score of the five clinical measures during the first three months of the trial (short-term effect) and during the final nine months of the trial (long-term effect). Paired *t*-tests were initially used to identify significant changes in scores. The changes were also evaluated using the standardized response mean (SRM) effect size statistic calculated as the ratio of the mean change of index scores to the standard deviation of the change scores [61]. The values 0.2, 0.5, and 0.8 were used as thresholds for small, moderate and large SRM effect sizes [62]. *T*-tests were used to identify significant differences between groups. Alongside parametric testing, non-parametric tests were explored. Finally, floor and ceiling effects were calculated as the proportion of participants at the lowest and highest level of capability and health respectively.

A priori hypotheses on the construct and discriminative validity of the measures are shown in the online Tables DS1 and DS2. Expectations about their sensitivity to change dictated that the ICECAP-A would be more sensitive to indicators related to broader wellbeing, with EQ-5D-5L being more sensitive to changes in physical health (TOP) and symptoms (CORE-OM).

Results

The trial recruited 83 participants, with 9 (11%) and 13 (15.6%) participants being lost to follow-up at 3 and 12 months post-randomisation. Participants had a mean age of 37 years, and were mostly male (86.8%) and single (90.4%). Most participants were also not in paid employment (80.7%), were receiving social security benefits (81.9%), and were in opiate substitution treatment for at least two years (79.5%). Mean capability (ICECAP-A) index score was 0.66 (SD = 0.19) and mean health (EQ-5D-5L) index score was 0.81 (SD = 0.20). More information about the sample characteristics is available in the online Table DS3.

Construct validity

The convergence of the ICECAP-A and EQ-5D-5L with the five clinical measures is shown in the correlation coefficient matrix of Table 1. Strong correlations ($r = 0.55$) were found between the ICECAP-A index score and the TOP dimensions of psychological health and overall quality of life. The correlation with the TOP dimension of physical health was statistically significant ($p < 0.01$) and in the moderate range. Strong correlations were also found between the ICECAP-A index score and the CORE-OM dimensions of wellbeing, symptoms, and functioning. The convergence with the risk dimension of the CORE-OM was marginally below the strong range ($r = -0.48$). Statistically significant correlations in the moderate range were found with LDQ ($r = -0.48$), SSQ ($r = 0.43$), and the ISEL dimensions of appraisal ($r = 0.36$) and self-esteem ($r = 0.30$).

In terms of the convergence between the EQ-5D-5L index score and the three TOP dimensions, a strong correlation ($r = 0.55$) was found with the dimension of psychological health, and a moderate correlation with the physical health ($r = 0.45$) and quality of life ($r = 0.34$) dimensions. A strong correlation was found between the EQ-5D-5L index score and the CORE-OM symptoms dimension, with the remaining correlations between the two measures being in the upper end of the moderate range. A low convergence with the ISEL was evident, as the EQ-5D-5L index score was not significantly correlated with the tangible, self-esteem, and belonging dimensions of the measure. Moderate to strong correlations were found between the EQ-5D-5L index score and the LDQ ($r = -0.48$) and SSQ ($r = 0.42$) scores. Using the EQ-5D-5L England value set, correlations with the dimensions of the five clinical measures were slightly higher compared with the use of the EQ-5D-3L UK crosswalk value set, but of similar strength (online Table DS5_A).

Expected correlations between the generic preference-based measures (ICECAP-A and EQ-5D-5L) and the five clinical measures were identified except for correlations involving the belonging dimension of the ISEL measure. The only hypothesised correlation that was not identified related to the convergence between the EQ-5D-5L index score and the ISEL self-esteem dimension ($r = 0.09$).

(Please insert Table 1 here)

Table 2 reports the result of the different analyses on discriminative validity. The findings for both the ICECAP-A and EQ-5D-5L confirmed all *a priori* hypotheses. However, some differences between the two measures need to be highlighted. Based on the results of the multivariate analysis, the ICECAP-A identified significant differences in capability-wellbeing, at the 1% level, across most dimensions of the five clinical measures. The only dimensions where differences in capability-wellbeing were statistically significant at the 5% level were for the dimensions of physical health status (TOP), belonging (ISEL), and social satisfaction (SSQ). For the former dimension, significant differences in HrQoL at the 5% level were identified by the EQ-5D-5L. For the latter two dimensions, however, differences in HrQoL were not statistically significant. This was also the case for the ISEL dimensions of appraisal and self-esteem. Using the EQ-5D-3L UK crosswalk value set in the supplementary analysis, the measure did not also identify statistically significant differences in the TOP dimensions of physical health status and overall quality of life as well as the CORE-OM risk dimension (online Table DS5_B).

(Please insert Table 2 here)

Sensitivity to change

There were no participants reporting the lowest level of capability or health across all five attributes of the ICECAP-A and EQ-5D-5L measures. There were, however, 22 (26.5%) participants reporting full health across all attributes of the EQ-5D-5L, while only 1 (1.2%) participant reported full capability. As evident in online Tables DS4_{A,B}, the proportion of participants at the highest level of capability ranged between 9.6% (*stability*) and 27.7% (*autonomy*), and the proportion of participants at the highest level of health ranged between 34.9% (*anxiety/depression*) and 89.2% (*self-care*). The distribution of the ICECAP-A responses for the 22 participants reporting full health based on the EQ-5D-5L, indicated that only 13.6%-27.3% of them were in a full capability in any of the five ICECAP-A attributes (online Table DS4_C).

Table 3 reports the sensitivity of the ICECAP-A and EQ-5D-5L index scores to changes in the number of days of abstinence from heroin, and the five clinical measures from baseline to 3-month follow-up. For participants with improved psychological health, wellbeing, self-esteem, belonging, dependence, and social satisfaction scores, changes in the ICECAP-A index score were statistically significant at the 1% level, with SRM effect sizes being approximately in the moderate range. Significant improvements in the ICECAP-A index score at the 5% level were also evident for those with more days of abstinence from heroin, improved physical health, overall quality of life, tangible support, and appraisal as well as for those without improved functioning and risk. For these, SRM effect sizes were in the small to moderate range. Statistically significant changes in the EQ-5D-5L score occurred only for those with improved levels of symptoms and social satisfaction (Table 3). For these, SRM effect sizes were moderate, while for most of the remaining improvements SRM effect sizes were mainly weak.

(Please insert Table 3 here)

Table 4 reports the sensitivity to change of the two measures between the 3 months and 12 months follow-up period. For those reporting improvements in psychological health, overall quality of life, functioning, appraisal, and social satisfaction, changes in the ICECAP-A index score were statistically significant, with the former being significant at the 1% level, and with SRM statistics ranging from small to large. Apart from these, many other changes were found with small SRM effect sizes. No statistically significant changes in the EQ-5D-5L index score were evident, and only improvements in symptoms, functioning, and self-esteem had weak SRM effect sizes. Statistically significant differences in capability-wellbeing were observed between those participants who improved overall quality of life, symptoms, functioning, appraisal, and social satisfaction, and those who did not. No significant differences in the EQ-5D-5L index score were found between the two groups. For both follow-up periods of the trial, the sensitivity of the EQ-5D-5L to changes in the dimensions of the clinical measures was similar regardless of the value set used (online Tables DS5_C and DS5_D).

(Please insert Table 4 here)

Discussion

This paper assessed the measurement properties of the ICECAP-A and EQ-5D-5L in a sample of heroin users receiving opiate substitution treatment. The analysis of construct validity indicated that both measures had the same level of convergence with the clinical measures, except when dimensions of broader wellbeing were involved, where the ICECAP-A showed a higher level of convergence. More specifically, correlations between the two measures and the dimensions of psychological and physical health, symptoms, risk, dependence and social satisfaction were of the same magnitude, while correlations with the dimensions of overall quality of life, wellbeing, functioning, and social support (appraisal and self-esteem) were stronger for the ICECAP-A.

The analysis of discriminative validity indicated that the ICECAP-A has as good discriminative properties as EQ-5D-5L for health-related constructs, and that it is able to capture broader wellbeing impacts that are likely to be missed by the EQ-5D-5L. These impacts were particularly evident in terms of the capacity to benefit from having close people to discuss personal issues and to socialise with or having personal self-esteem.

The sensitivity to change analysis explored changes in HrQoL and capability-wellbeing in response to changes in a number of different clinical anchors at 3 and 12 months post-randomization. The results of the analysis indicated that for 11 out of the 14 anchors used in the short-term follow-up, participants reporting clinical improvements had a statistically significant increase in capability-wellbeing, while in only 2 anchors there were statistically significant improvements in HrQoL.

Nevertheless, in 10 out of the 14 anchors the SRM statistic for the EQ-5D-5L was above the lower threshold of weak effect sizes (0.2). Although this may provide some supporting evidence on the sensitivity to change for the EQ-5D-5L, the findings from the long-term follow-up indicate low sensitivity. Improvements with at least a small SRM effect size were reported in only 3 out of 14 EQ-

5D-5L anchors, compared with the same change in 9 anchors with the ICECAP-A. Similarly, while participants reporting deteriorations in 1 anchor had at least small SRM for the EQ-5D-5L, at least small SRM effect sizes were found in 6 anchors for the ICECAP-A. The use of the EQ-5D-5L England instead of the EQ-5D-3L UK crosswalk value set appears to have slightly improved the psychometric performance of the measure. The EQ-5D-5L England value set led to much higher HrQoL estimates and with lower variability.

Several reasons could potentially explain the psychometric performance of the ICECAP-A and EQ-5D-5L in this clinical context. Firstly, the descriptive system of the ICECAP-A incorporates more determinants of wellbeing than health status. This explains why the ICECAP-A showed strong psychometric properties, not only in terms of overall quality of life, wellbeing, functioning, social support and satisfaction but also in terms of physical and psychological health, and symptoms. Secondly, the sensitivity of the EQ-5D-5L was hindered by a ceiling effect. In the study, 22 (26.5%) participants reported full health, with only 1 (1.2%) being in full capability. The ceiling effect was particularly evident in the *self-care* (89.2%), *mobility* (74.7%), and *usual activities* (72.3%) dimensions, which raises concerns about their relevance in this context.

To the best of our knowledge, this is the first study exploring the psychometric properties of the ICECAP-A and EQ-5D-5L in the context of substance use disorders. However, some of the findings for the EQ-5D-5L are similar to the ones from other studies on EQ-5D-3L. Research has shown that improvements in participants receiving opiate substitution treatment are commonly evident in the dimensions of *pain/discomfort* (33.2%), *anxiety/depression* (31.5%), and *usual activities* (30.3%) [36]. These dimensions were also the most responsive in this study with 43.2%, 41.9%, and 32.4% of participants reporting improvements in the three dimensions respectively during the first three months of the trial. This is possibly due to the large proportion of participants clustered in the full health option in the *mobility* and *self-care* dimensions. Validation work on the EQ-5D-3L in a heroin-dependent population has indicated that 91.4% and 81.9% of participants reported full health in terms of *self-care* and *mobility*, with the relevant proportion in the rest of the attributes ranging between

51.6% (*pain/discomfort*) and 62.2% (*usual activities*) [37]. Considering the large ceiling effect found in this study, it appears that the extension of response options to the EQ-5D-3L did little to improve the sensitivity of the measure, and possibly that the conceptual attributes of the measure do not comprehensively capture important patient outcomes in this context, as has recently been reported for the wider mental health context [63].

The study targeted participants in opiate substitution treatment who still reported heroin use, as they represent a group that might be perceived to have made sub-optimal progress in treatment. The analysis of short- and long-term follow-up data and the use of five clinical measures constitute major strengths of this study, which allowed for a comprehensive assessment of the psychometric properties of the two measures. There are, however, a number of limitations worth mentioning. First, the study relied on a relatively modest sample size. Second, in the absence of a gold-standard measure of wellbeing, the psychometric properties of the two measures could only be explored based on hypothetically developed constructs from specific demographic or contextual variables and other clinical measures. Third, the analysis explored the sensitivity of the two measures on the basis of overall changes in different clinical dimensions and not on the basis of clinically meaningful changes, which concerns economic evaluations [61]. Finally, the findings of this study are restricted to the treatment of opiate dependent patients, and thus transferability of conclusions requires caution.

Establishing the psychometric performance of a measure is a continuous process and more research is needed to explore how well the ICECAP-A and EQ-5D-5L perform in different social care contexts. Comparisons with other capability measures developed for social care (ASCOT) [64] and mental health (OxCap-MH) [65] or with other measures of HrQoL will contribute toward gaining further understanding of their psychometric properties. Previous evidence has indicated that a shift from HrQoL to capability-wellbeing would prioritise funding for those suffering from depression and other severe illnesses [66]. Based on the findings of this study, this is also likely to be the case for the context of drug addiction, where treatment objectives extend beyond the improvement of health. The ICECAP-A was found to be at least as sensitive as EQ-5D-5L for health-related clinical indicators and

more sensitive for clinical indicators related to broader wellbeing, and therefore its use in economic evaluations in this context is recommended. Due to the role of the EQ-5D in current decision making process, both measures should be used in the economic analysis of interventions in the context of substance use disorders, in line with NICE's social care recommendation.

REFERENCES

- [1] National Institute for Health and Care Excellence (NICE). Guide to the methods of technology appraisal 2013. London: NICE, 2013.
- [2] Drummond M, Sculpher M, Torrance G, et al. Methods for the economic evaluation of health care programmes. New York: Oxford University Press, 2005.
- [3] Brooks R, EuroQol Group. EuroQol: the current state of play. *Health policy*. 1996; 37: 53-72.
- [4] Brazier J, Ratcliffe J, Salomon J, et al. Measuring and valuing health benefits for economic evaluation. New York: Oxford University Press, 2007.
- [5] Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of life research*. 2011; 20: 1727-36.
- [6] Conner-Spady BL, Marshall DA, Bohm E, et al. Reliability and validity of the EQ-5D-5L compared to the EQ-5D-3L in patients with osteoarthritis referred for hip and knee replacement. *Quality of Life Research*. 2015: 1-10.
- [7] Golicki D, Niewada M, Buczek J, et al. Validity of EQ-5D-5L in stroke. *Quality of Life Research*. 2014: 1-6.
- [8] Greene ME, Rader KA, Garellick G, et al. The EQ-5D-5L Improves on the EQ-5D-3L for Health-related Quality-of-life Assessment in Patients Undergoing Total Hip Arthroplasty. *Clinical Orthopaedics and Related Research*. 2014: 1-8.
- [9] 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. *Quality of Life Research*. 2013; 22: 1717-27.
- [10] Tordrup D, Mossman J, Kanavos P. Responsiveness of the EQ-5D to clinical change: is the patient experience adequately represented? *International journal of technology assessment in health care*. 2014; 30: 10-19.
- [11] Longworth L, Yang Y, Young T, et al. Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. *Health Technol Assess*. 2014; 18.
- [12] Brazier J, Connell J, Papaioannou D, et al. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. *Health Technol Assess*. 2014; 18.
- [13] Al-Janabi H, Flynn TN, Coast J. QALYs and carers. *Pharmacoeconomics*. 2011; 29: 1015-23.
- [14] Chalkidou K, Culyer A, Naidoo B, et al. Cost-effective public health guidance: asking questions from the decision-maker's viewpoint. *Health Economics*. 2008; 17: 441-48.
- [15] Coast J. Strategies for the economic evaluation of end-of-life care: making a case for the capability approach. *Expert Review of Pharmacoeconomics & Outcomes Research*. 2014; 14: 473-82.
- [16] Lorgelly PK, Lawson KD, Fenwick EA, et al. Outcome measurement in economic evaluations of public health interventions: a role for the capability approach? *International Journal of Environmental Research and Public Health*. 2010; 7: 2274-89.
- [17] Makai P, Brouwer WB, Koopmanschap MA, et al. Quality of life instruments for economic evaluations in health and social care for older people: A systematic review. *Social Science & Medicine*. 2014; 102: 83-93.

- [18] Payne K, McAllister M, Davies LM. Valuing the economic benefits of complex interventions: when maximising health is not sufficient. *Health economics*. 2013; 22: 258-71.
- [19] Knapp M, Mangalore R. The trouble with QALYs.... *Epidemiologia e psichiatria sociale*. 2007; 16: 289-93.
- [20] Sen A. Capability and well-being. In: Nussbaum M, Sen A, eds., *The quality of life*. Oxford: Oxford University Press, 1993.
- [21] Coast J, Flynn TN, Natarajan L, et al. Valuing the ICECAP capability index for older people. *Social science & medicine*. 2008; 67: 874-82.
- [22] Al-Janabi H, Flynn TN, Coast J. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Quality of Life Research*. 2012; 21: 167-76.
- [23] Sutton EJ, Coast J. Development of a supportive care measure for economic evaluation of end-of-life care using qualitative methods. *Palliative medicine*. 2014; 28: 151-57.
- [24] Darke S. *The Life of the Heroin User: Typical Beginnings, Trajectories and Outcomes*. Cambridge, UK: Cambridge University Press, 2011.
- [25] Sindelar JL, Jofre-Bonet M, French MT, et al. Cost-effectiveness analysis of addiction treatment: paradoxes of multiple outcomes. *Drug and alcohol dependence*. 2004; 73: 41-50.
- [26] De Maeyer J, Vanderplasschen W, Broekaert E. Quality of life among opiate-dependent individuals: A review of the literature. *International Journal of Drug Policy*. 2010; 21: 364-80.
- [27] Uchtenhagen A. Abstinence versus Agonist Maintenance Treatment: An Outdated Debate? *European Addiction Research*. 2013; 19: 283-6.
- [28] Department of Health. *Drug strategy 2010 Reducing demand, restricting supply, building recovery: Supporting people to live a drug free life*. London: Department of Health, 2010.
- [29] UK Drug Policy Commission (UKDPC). *The UK Drug Policy Commission recovery consensus group: a vision of recovery*. London: UKDPC. 2008.
- [30] Scottish Government. *The Road To Recovery: A New Approach To Tackling Scotland's Drug Problem*. Scottish Government, 2008.
- [31] National Institute for Health and Care Excellence (NICE). *The social care guidance manual*. London: NICE, 2013.
- [32] Al-Janabi H, Peters TJ, Brazier J, et al. An investigation of the construct validity of the ICECAP-A capability measure. *Quality of Life Research*. 2013; 22: 1831-40.
- [33] 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. *Quality of Life Research*. 2016; DOI 10.1007/s11136-015-1225-y.
- [34] Keeley T, Al-Janabi H, Lorgelly P, et al. A qualitative assessment of the content validity of the ICECAP-A and EQ-5D-5L and their appropriateness for use in health research. *PloS one*. 2013; 8: e85287.
- [35] 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. *Quality of Life Research* DOI: 101007/s11136-015-0980-0. 2015.
- [36] Nosyk B, Guh DP, Sun H, et al. Health related quality of life trajectories of patients in opioid substitution treatment. *Drug and alcohol dependence*. 2011; 118: 259-64.
- [37] van der Zanden BP, Dijkgraaf MG, Blanken P, et al. Validity of the EQ-5D as a generic health outcome instrument in a heroin-dependent population. *Drug and alcohol dependence*. 2006; 82: 111-18.

- [38] Pyne JM, French M, McCollister K, et al. Preference-weighted health-related quality of life measures and substance use disorder severity. *Addiction*. 2008; 103: 1320-29.
- [39] Pyne JM, Tripathi S, French M, et al. Longitudinal association of preference-weighted health-related quality of life measures and substance use disorder outcomes. *Addiction*. 2011; 106: 507-15.
- [40] Byford S, Barrett B, Metrebian N, et al. Cost-effectiveness of injectable opioid treatment v. oral methadone for chronic heroin addiction. *The British Journal of Psychiatry*. 2013; 203: 341-49.
- [41] Dijkgraaf MG, van der Zanden BP, de Borgie CA, et al. Cost utility analysis of co-prescribed heroin compared with methadone maintenance treatment in heroin addicts in two randomised trials. *Bmj*. 2005; 330: 1297.
- [42] Day E, Copello A, Seddon JL, et al. Pilot study of a social network intervention for heroin users in opiate substitution treatment: study protocol for a randomized controlled trial. *Trials*. 2013; 14: 264.
- [43] Copello A, Orford J, Hodgson R, et al. Social behaviour and network therapy - basic principles and early experiences. *Addictive Behaviours*. 2002; 27: 345-66.
- [44] Copello A, Williamson E, Orford J, et al. Implementing and evaluating Social Behaviour and Network Therapy in drug treatment practice in the UK: a feasibility study. *Addictive Behaviors*. 2006; 31: 802-10.
- [45] Day E, Best D, Bartholomew N, et al. The BTEI Care Planning Manual: Mapping Achievable Goals. In: *Routes to Recovery*. National Treatment Agency for Substance Misuse, 2008.
- [46] Marsden J, Farrell M, Bradbury C, et al. Development of the treatment outcomes profile. *Addiction*. 2008; 103: 1450-60.
- [47] Evans JM-C, Frank Margison, Michael Barkham, Kerry Audin, Janice Connell, Graeme McGrath, Chris. CORE: Clinical outcomes in routine evaluation. *Journal of Mental Health*. 2000; 9: 247-55.
- [48] Cohen S, Hoberman HM. Positive events and social supports as buffers of life change stress. *Journal of applied social psychology*. 1983; 13: 99-125.
- [49] Raistrick D, Bradshaw J, Tober G, et al. Development of the Leeds Dependence Questionnaire (LDQ): a questionnaire to measure alcohol and opiate dependence in the context of a treatment evaluation package. *Addiction*. 1994; 89: 563-72.
- [50] Tober G, Brearley R, Kenyon R, et al. Measuring outcomes in a health service addiction clinic. *Addiction Research & Theory*. 2000; 8: 169-82.
- [51] Barkham M, Margison F, Leach C, et al. Service profiling and outcomes benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology*. 2001; 69: 184.
- [52] Brookings JB, Bolton B. Confirmatory factor analysis of the interpersonal support evaluation list. *American journal of community psychology*. 1988; 16: 137-47.
- [53] Raistrick D, Tober G, Heather N, et al. Validation of the Social Satisfaction Questionnaire for outcome evaluation in substance use disorders. *Psychiatric Bulletin*. 2007; 31: 333-36.
- [54] Flynn TN, Huynh E, Peters TJ, et al. Scoring the ICECAP-A capability instrument. Estimation of a UK general population tariff. *Health economics*. 2015; 24: 258-69.
- [55] Devlin N, Shah K, Feng Y, et al. Valuing Health-Related Quality of Life: An EQ-5D-5L Value Set for England. London: Office of Health Economics, 2016.
- [56] Dolan P. Modeling valuations for EuroQol health states. *Medical care*. 1997; 35: 1095-108.
- [57] van Hout B, Janssen M, Feng Y-S, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value in Health*. 2012; 15: 708-15.

- [58] Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. New York: Oxford University Press, 2003.
- [59] Cohen J. Set correlation and contingency tables. *Applied Psychological Measurement*. 1988; 12: 425-34.
- [60] Raistrick D, Tober G, Sweetman J, et al. Measuring clinically significant outcomes-LDQ, CORE-10 and SSQ as dimension measures of addiction. *Psychiatric Bulletin*. 2014; 38: 112-15.
- [61] Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: Learning from psychometrics. *Health Economics*. 1999; 8: 41-51.
- [62] Cohen J. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum Associates, 1988.
- [63] Connell J, O'Cathain A, Brazier J. Measuring quality of life in mental health: Are we asking the right questions? *Social Science & Medicine*. 2014; 120: 12-20.
- [64] Netten A, Burge P, Malley J, et al. Outcomes of social care for adults: developing a preference-weighted measure. *Health Technology Assessment*. 2012; 16: 1-166.
- [65] Simon J, Anand P, Gray A, et al. Operationalising the capability approach for outcome measurement in mental health research. *Social Science & Medicine*. 2013; 98: 187-96.
- [66] Mitchell P, Al-Janabi H, Richardson J, et al. The Relative Impacts of Disease on Health Status and Capability Wellbeing: A Multi-Country Study. *PLoS ONE*. 2015; 10: e0143590. doi:10.1371/journal.pone.0143590.