

WORD COUNT: 7978

# **VALUING THE ICECAP CAPABILITY INDEX FOR OLDER PEOPLE**

**Joanna Coast**

*Department of Health Economics, University of Birmingham*

**Terry N Flynn**

*Department of Social Medicine,, University of Bristol*

**Lucy Natarajan**

*National Centre for Social Research*

**Kerry Sproston**

*National Centre for Social Research*

**Jane Lewis**

*National Centre for Social Research*

**Jordan J Louviere**

*Centre for the Study of Choice, University of Technology, Sydney*

**Tim J Peters**

*Department of Community Based Medicine, University of Bristol*

Address for correspondence:  
Joanna Coast  
Professor of Health Economics  
Department of Health Economics  
Public Health Building  
University of Birmingham  
Birmingham  
B15 2TT  
Tel: 0121 414 3056  
Fax: 0121 414 8969  
E-mail: [j.coast@bham.ac.uk](mailto:j.coast@bham.ac.uk)

## SUMMARY

This paper reports the first application of the capabilities approach to the development and valuation of an instrument for use in the economic evaluation of health and social care interventions. The ICECAP index of capability for older people focuses on quality of life rather than health or other influences on quality of life, and is intended to be used in decision making across health and social care in the UK. The measure draws on previous qualitative work in which five conceptual attributes were developed: attachment, security, role, enjoyment and control. This paper details the innovative use within health economics of further iterative qualitative work among 19 informants to refine lay terminology for each of the attributes and levels of attributes used in the eventual index. For the first time within quality of life measurement for economic evaluation, a best-worst scaling exercise has been used to estimate general population values (albeit for population of those aged 65+) for the levels of attributes, with values anchored at 1 for full capability and 0 for no capability. Death was assumed to be a state in which there is no capability. The values obtained indicate that attachment is the attribute with greatest impact but all attributes contribute to the total estimation of capability. Values that were estimated are feasible for use in practical applications of the index to measure the impact of health and social care interventions.

# INTRODUCTION

Current UK policy regarding the provision of both health care and social care for older people suggests that greater integration is required between these two areas (Glendinning, 2003; Department of Health, 2001). Economists' attempts to assist such resource allocation decisions, however, strongly focus on measuring health (in its broadest sense), with proxies for health, life expectancy, and health-related quality of life measures (in particular the Quality-Adjusted Life Year (QALY) (Williams, 1985)) dominating the empirical economic evaluation literature. Many social services interventions, however, may impact more broadly on quality of life (assumed here to encompass the broad range of factors that are important to people in living their lives) rather than health (assumed here to include aspects of physical and mental health). Measures that look only at health in assessing the impact of these interventions would be very likely to underestimate impact.

Given the concern for closer integration, previous work by this research team has concentrated on developing attributes for a new measure of (general) quality of life for older people, rather than health-related quality of life or influences on quality of life (Grewal, Lewis, Flynn, Brown, Bond & Coast, 2006). The purpose of such a measure would be to provide information for decisions about the allocation of resources across health and social care, rather than just across health. By including dimensions that are concerned with quality of life rather than health alone the measure would enable comparisons across a broader range of interventions. For example, the measure would more easily enable decision makers to compare the value of social service interventions that may improve quality of life without improving health (such as: aids and adaptations including, for example, housing interventions; day care; and meals on wheels) with, for example, health interventions that improve both health and

quality of life (such as provision of drugs for Alzheimer's or the surgical replacement of hips and knees). This type of measure might also be extremely useful for interventions concerned with the public health of older people, such as mental health services for dementia and depression, measures to reduce falls, continence services, etc. (Department of Health & OPD(PIP), 2004).

Prior work by the research team used in-depth interviews with older people to find out what mattered to them in terms of their quality of life (Grewal et al., 2006). Although discussion initially concentrated upon factors influencing quality of life (activities, relationships, health, wealth, surroundings and religion/fait h/spirituality) further probing and analysis suggested that five conceptual attributes were important: attachment (feelings of love, friendship, affection and companionship); role (the idea of having a purpose or "doing something" that is valued, either by the individual and/or by others); enjoyment (notions of pleasure and joy, and a sense of satisfaction); security (feeling safe and secure, not having to worry and not feeling vulnerable); and control (being independent and able to make one's own decisions) (Grewal et al., 2006).

Importantly, the previous study also suggested that informants' quality of life was limited by loss in ability to pursue these attributes (Grewal et al., 2006). So, for example, poor health itself did not reduce quality of life; rather, the important issue was the influence such poor health had upon informants' abilities to achieve these attributes. The work was then linked with the extensive literature on capabilities (Robeyns, 2003; Robeyns, 2005; Nussbaum, 2003) associated particularly with the work of Amartya Sen (Sen, 1992; Sen, 1993; Sen, 1982; Sen, 2002). Sen distinguishes between functioning and capability as possible foci for evaluation, using the example of the person who is starving due to lack of food compared with the person for whom food is freely available, but who chooses to fast, to indicate why capability instead of functioning might be an important focus of evaluation (Sen, 1993). Here, the

five conceptual attributes developed in the previous work are interpreted as a set of functionings, the capability to achieve which appeared to be of importance. The previous work concluded that further development of the measure would focus on developing an index of capability (Grewal et al., 2006). Such an approach has not previously been pursued either within the capabilities or the health economics literature.

It should be stressed that developing a capability index was not the starting point for the original research, but rather one that emerged from careful analysis of the qualitative data. Consequently, whilst an index of capability seems to most closely reflect the values of the older people interviewed and thus draws to a greater extent on the capabilities literature than previous “extra-welfarist” assessments within health economics (Culyer, 1990; Culyer, 1989), the research is consistent with the economic evaluation paradigm in health care research, with the measure intended to provide a useful outcome for economic evaluation. To be useful for this purpose, the measure must be able to be completed by older people participating in studies to determine the impact of interventions and must be linked to a set of values that can provide a weighted measure of outcome. The standard approach within health economics, however, is to use preference elicitation methods that weight (anchor) health against death but here a different approach has been used to anchor the index.

This paper details a number of innovative methods that have been used to develop the measure of capability (the ICECAP\* measure) from the point at which the five conceptual attributes were obtained. In particular, the application of qualitative methods within health economics to develop terminology for the measure, the use of best-worst scaling to develop capability values and anchoring in terms of capability rather than death, represent clear

---

\* ICECAP stands for ICEpop CAPability index. ICEPOP is the name of the UK MRC-funded programme through which the index was developed.

departures from usual practice within the development of measures for use in economic evaluation. The paper thus not only presents results for the final terminology used in the index and the values obtained from a general population sample of older people but illustrates the successful application of these methods in the context of health economics.



# METHODS

## Qualitative work: design

The attributes that emerged from the initial qualitative work were attachment, role, enjoyment, control and security (Grewal et al., 2006). These conceptual terms were clearly not described in ways that would be meaningful to older people who are asked to complete a measure nor to policy makers who have to interpret results. In terms of using an index based on these attributes for policy evaluation in health and social care, those completing the index need to be able to understand the meaning behind the conceptual labels so that they can indicate the extent to which an attribute is attainable in their lives. In contrast, the work by Anand and van Hees which essentially takes the same approach of asking people about their capabilities uses academic terminology in its classification system such as “intellectual stimulation” and “social relations” (Anand & van Hees, 2006) – this is not necessarily meaningful to people and might therefore evoke different meanings from those intended.

In moving from conceptual attributes to a classification system that can be used with older people, three steps were necessary. First, it was necessary to ensure that the terminology used represented the conceptual attributes, but expressed them in language meaningful to older people completing the classification system. Second, it was necessary to ensure that the terminology for the different attribute levels represented the notion of capability rather than functioning. Third, it was necessary to develop levels for each attribute using appropriate and meaningful terminology.

A further phase of qualitative work with older people was therefore conducted. Semi-structured interviews were conducted with 19 informants interviewed

previously. Informants aged 65 and over were selected from respondents to a general population survey, the National Survey of NHS Patients (Boreham, Airey, Erens & Tobin, 2002). Sampling for the initial survey was restricted to three locations and was purposive to include the range of personal characteristics: sex, age, health status, household composition, and current/most recent occupation. For this second phase informants were approached again and asked for consent to take part in a second interview. Those agreeing were visited in their own homes where interviews were tape recorded.

This phase of the research used iterative techniques to constantly refine language for both attributes and levels. The process began with the research team brainstorming for lay terms for each concept. These terms were then explored in interviews, with informants being asked to express what the terms meant to them. All interviews were transcribed verbatim. Analysis, using constant comparison (Glaser & Strauss, 1968; Strauss & Corbin, 1990) and the writing of accounts to clarify and illuminate different areas, was iterative and ongoing throughout the research. For each iteration a small number of interviews was conducted and analysed, and changes were made to the interview schedule before conducting the next set of interviews. When saturation was achieved for particular topics (that is, the meanings of terms was clear to respondents), these topics were not considered further. During successive interviews terms were tested, altered and refined until the final terminology for each attribute was achieved. During the final set of interviews terms were reworded as capability levels to obtain views about the meanings of different attribute levels and the use of the capability terminology.

## Quantitative work: valuation

The definitive process by which values should be elicited for capabilities is unresolved. It is clear that Sen “explicitly rejects the use of either choices or desires to value capabilities” (Cookson, 2005) and that he is concerned that the capability approach should not rest exclusively on mental states like happiness, with moral judgements also needing to be incorporated into a capability index (Robeyns, 2005). Much of the capability approach is concerned with democratic debate and deliberation – particularly for the development of “lists” of capabilities, but also in relation to their values and weights. Indeed, in a recent paper Sen refers to “the need for open valuational scrutiny for making social judgements” (Sen, 2005). It is not clear, however, how “processes of public reasoning and democracy are going to take place, and how we can make sure that minimal conditions of fair representation are guaranteed” (Robeyns, 2005). Practical issues associated with eliciting values from people via such a process are not discussed in the capabilities literature.

Cookson has suggested one way around the problem of reconciling the capabilities approach with extra-welfarist health economics, which is to think of the views elicited from populations as value judgements, rather than preferences (Cookson, 2005). In this sense, preferences are seen as an individual’s choices or desires which in part depend on their expectations, whereas value judgements are seen as being “normative, and by nature susceptible to reasoned assessment.” (Cookson, 2005), p.822) Using Cookson’s approach would imply that it is possible to obtain information from the relevant population about their value judgements, and to use this as evidence for the choice of values given to each dimension (Cookson, 2005). Such use would also be somewhat consistent with riskless value functions, but inconsistent with expected utility theory (Keeney & Raiffa, 1976). Essentially, then, the values

obtained from a population would provide evidence regarding the final value in the capability index but would not necessarily constrain them.

## **Experimental design**

The approach adopted here has been to elicit a set of values for initial use in the capability index from a population of older people. The best-worst scaling method has been used along with a stated preference discrete choice experiment. Best-worst scaling is a form of stated preference discrete choice modelling (that is, people are asked to make choices, and these choices reveal their values for particular attributes) (Flynn, Louviere, Peters & Coast, 2007; Marley & Louviere, 2005; McIntosh & Louviere, 2002). In the profile (or 'attribute') version of best-worst scaling used here, respondents are presented with a set of hypothetical scenarios (Marley, Flynn & Louviere, 2008). Each scenario offers respondents a combination of attribute levels and asks them to choose which one of the attribute levels is the best (most attractive) and which one of the attribute levels is the worst (least attractive); respondents are essentially asked to choose that pair of attributes that maximises the difference in value between them. Scenarios are constructed by combining the levels of each attribute using an experimental design.

Although best-worst scaling can be viewed as a choice experiment, respondents are asked only to specify which attribute levels they think are the best and the worst. Hence, it can be argued that values and not preferences are elicited, because individuals are not asked to trade one thing for another. Thus, the best-worst scaling approach may come closer to eliciting values that would satisfy Sen's interpretation than measures produced by other methods advocated by economists. Best-worst scaling has other potential advantages over traditional discrete choice experiments: first, because it offers more insights into preference heterogeneity than 'pick one' tasks; and second, because it is potentially less cognitively demanding.

Given the overall classification system size of five attributes, each with four levels, there were 1024 possible scenarios which respondents could have been given in the valuation task. Providing all scenarios in a single task was not feasible and hence the number of scenarios was reduced by designing one version based on the orthogonal main effects plan (OMEP), with a second statistically equivalent version constructed from what is known as the “foldover” of the first version. Specifically, an OMEP was used to create 16 scenarios. The OMEP was obtained from the website <http://www.research.att.com/~njas/oadir/oa.16.5.4.2.txt>. This OMEP gives a set of scenarios in which all attributes are statistically independent of one another, making it possible to independently estimate the values that people associate with each level of every attribute provided that all interactions between attributes are zero. Because 16 scenarios is relatively few compared to the number of parameters of interest, a second set of 16 scenarios using the foldover of the OMEP was also constructed, giving 32 scenarios in total. Respondents were randomly assigned to one of two groups, with each group receiving one of the two versions of 16 scenarios. Group A received 16 scenarios constructed from the original OMEP and group B were given 16 scenarios constructed from the foldover which gives a statistically equivalent design. Attributes and scenarios were ordered randomly in each version.

## **Sampling**

The sample was obtained using the sampling frame of respondents to the Health Survey for England (HSE) which is conducted on the basis of postcode sectors and using a general population sample drawn from the Postcode Address File (publicly available) and stratified by the percentage of non-manual workers in the postcode sector and by Local Authority. To be eligible to take part in the valuation survey respondents had to have been previously interviewed as part of HSE 2005, to have agreed to be re-interviewed and to be

aged 65 and over. Clusters of sectors were randomly selected from within three regional fieldwork areas covering the north to the south of England, with all eligible people in the selected areas being included.

### **Sample size**

Lack of necessary information on, for instance, variability, precluded traditional power calculations to determine the survey sample size which, instead, was determined using simulations. These demonstrated that, for a design with five attributes, four levels per attribute, and 16 scenarios per respondent, 100 completed questionnaires would be adequate to estimate the underlying values (for a wide range of likely variances such that an R-squared of at least 90% would be achieved when regressing estimated values against the true values). It was assumed in these simulations, however, that all individuals had similar values. One aim of this work (not reported here) was to explore the extent to which values vary and so the desired sample size was increased to 300 completed questionnaires. A response rate of 65% was assumed and so the minimum number of households to be approached was 460, requiring 92 postcode sectors to be sampled.

### **Data collection**

A structured survey schedule was used in which cards in large type size were used to present the scenarios one-by-one to respondents. Each card presented one scenario. Each scenario described a particular capability state, specified in terms of particular levels of the five attributes, so that, for example, the top level of attachment might be combined with the top level of security, the bottom level of role, and middle levels from each of enjoyment and control. From such a scenario the respondent was asked to select what aspect of that scenario would be the best thing about it and what would be the worst. Respondents were then asked supplementary questions about whether they would prefer the described

state to their own health state and, if not, whether they felt that the described state was a life worth living.

Data were collected by interviewers who had attended a full day briefing. Interviews were held in the respondents' homes. Prior to full data collection the survey schedule was piloted with 30 respondents. Of particular concern was whether respondents would be able to cope with 16 scenarios, and the pilot was confirmed that this was the case.

Information from completed questionnaires was input initially into SPSS.

### **Data analysis**

Best-worst data can be analysed in a number of different ways (Flynn et al., 2007). Here, values (part-worth utilities) were estimated using multinomial (conditional) logistic regression. The unit of analysis in the paired model (Flynn et al., 2007) (also called 'maxdiff' (Marley & Louviere, 2005)) is the best-worst pair of attribute levels chosen in a scenario.. Analysis was conducted for both the sub sample that provided only complete best-worst responses to all scenario questions and for the sub sample that provided any best-worst choice data, whether these responses were complete or not. This analysis suggested that the estimated values differed very little and, to maintain consistency with individual level analyses that will be presented elsewhere, data are presented here for the sub sample with complete best-worst choice data. Best-worst scaling produces estimates of values on an interval scale. These were initially scaled such that the lowest valued attribute level was given the value zero. Stata version 9SE was used for analysis (Stata Corporation, 2003); the clogit regression command was used to estimate part-worth utilities, with coding of the explanatory variables as detailed by Flynn et al. (2007).

## **Anchoring**

One aspect of the valuation of health outcomes in health economics is the desire to integrate information about quality and quantity of life. This is because interventions in the health field frequently impact on either quality or quantity of life, or both, and there is a desire to be able to compare across interventions that provide these different benefits (National Institute for Clinical Excellence, 2004). Fully integrating and comparing these dimensions raises issues about the meaning of anchoring at death as is done in the QALY and how measures in which death is included on the same scale as health/quality of life can be interpreted. For example, does the absence of life equal the absence of health? Or, in the case addressed in this paper, does the absence of life equal the absence of capability?

There are therefore questions about how to deal with potential mortality resulting from an intervention. At no point does the capabilities literature suggest that capabilities should be combined with length of life. Indeed, the identification of basic capabilities such as the ability to be well-nourished as important (Sen, 1992) suggests that the potential for mortality might be accounted for through a different mechanism and more indirectly than it would be in traditional health-related utility indices.

One option would be to use a preference-based measure to provide a value for death, against which all other values could be rescaled in a manner similar to the QALY. This would clearly be counter to Sen's views with regard to the use of preferences in valuation (see above), and also encounters the problem common to all QALY measurement – that people are having to value against death, with little idea of what death involves (Carr-Hill, 1989).

An alternative would be to anchor on a more philosophical basis, such that the absence of capability is given a value of zero. A number of states may produce such a zero value: assessment of capabilities as being non-existent in relation to



all attributes; unconsciousness; and death. This option has the benefit of being able to measure the capabilities of all states without needing to value states against death. However, it also enables death to be evaluated against and in terms of capabilities. Indeed, it is noted in a paper concerned with the Human Development Index that survival is a prerequisite for enjoying other capabilities (Anand & Sen, 1994). This option appears to accord more closely with Sen's ideas, by avoiding the use of preferences to evaluate death. For this reason, this alternative approach is adopted here.

To anchor the values, they were rescaled such that the state of the "absence of capability" (state 11111, that is, the lowest capability state of level 1 - no capability - on each attribute) was given a total value of zero, and the state of full capability on all attributes (state 44444, that is, the highest capability state of level 4 - full capability - on each attribute) was given a value of one. This was achieved by, first, subtracting one fifth of the index value of state 11111 from all attributes and then dividing by the index value for 44444. Such rescaling ensures that the individual capability attribute level ones ("1"s) sum to zero, that the full capability level fours ("4"s) sum to one and that the relative differences between attributes and attribute levels are maintained.

# RESULTS

## Qualitative work: design

Table 1 shows the characteristics of informants interviewed during the design stage for the measurement instrument. For reasons of space, it is not feasible to report the development of the terminology for the instrument in detail. Instead, the “security” attribute is used as an exemplar, as this was the most difficult term to clarify. This concept was defined originally by the research team as “ideas of feeling safe and secure, not having to worry and not feeling vulnerable”, influences upon which include having sufficient finances, sufficient practical and emotional support and sufficient health.” A number of different ways of expressing this concept were initially developed by the research team, including: security, not feeling vulnerable, having money or support so you don’t have to worry. Initial interviews indicated the following meanings for the term “vulnerable”:

*“... I see vulnerable as being if you’re out after dark... vulnerable to things that may or may not happen...” (Male, age 77)*

*“...vulnerable to me almost implies danger, somebody’s going to come along and cosh you on a dark night... But it’s not what we’re trying to say I don’t think. You’re trying to say vulnerable in the sense that you’re not going to be able to cope and you’re not going to be left not to cope...” (Male, age 76)*

This term was consequently abandoned because it clearly evoked the wrong meaning for at least some informants, as did the term “security” which for one early respondent evoked a focus on national security:

*"At the moment I'm looking at the blackest scene to think we're certainly going to have a huge conflagration throughout the world... We've got America that's deciding... I think of the bigger picture." (Male, age 70)*

An alternative terminology considered was "money or support so you don't have to worry". Again, these terms were not interpreted as intended because money was an immediate focus of informants to the exclusion of other aspects of security:

*"Just as long as I have enough to see me through..." (Female, age 73)*

*"...financial support to me is state benefits or something like that" (Female, age 66)*

The next term investigated was "not having to worry about things". This also evoked a concern only for finances for some people, and for others, the use of the word "worry" also conflicted with their own self image, making it difficult for them to identify with the terminology:

*"I am not a worrier, I've never been a worrier because I always think 'what's the good of worrying, it'll either come out or it won't'... so I think that is a lot to do with me individually." (Female, age 85)*

Adding a notion of the future seemed to deflect people's thoughts away from current monetary concerns, however, so the terminology "thinking about the future without having to worry" was tried, which seemed to evoke meanings closer to those originally envisaged:

*"... to know that you're going to be able to live reasonably securely, secure in health, financial security and probably emotional security" (Male, age 70)*

The use of the terminology of "worry" was still not completely evoking the right responses, however, and so the terminology finally settled on was "thinking about the future without concern":

*"I often think just briefly, 'Well who's going to look after me? 'Cos the family are too far away.'" (Female, age 74)*

Language used in the eventual measure was "love and friendship" (attachment), "doing things that make you feel valued" (role), "enjoyment and pleasure" (enjoyment), "thinking about the future without concern" (security), "being independent" (control). The attributes that were most difficult to clarify were security and control (where introducing the notion of making choices, made people focus exclusively on decision making and not other aspects of control).

In general, informants understood the capability terminology and the different meanings associated with different attribute levels:

*"... it would be that one, 'you are unable to do things [that make you feel valued]' ... If you couldn't go and see to somebody you that you knew and was sick and they needed you..." (Female, age 66)*

*"I can only think about the future with a lot of concern, and that is purely two things I put in that category. One, of course, is the most important, which is your health. And the other is finance. Because as time goes on, it's not getting any easier." (Female, age 69)*

Initially three levels were assigned to each attribute, but people appeared to have difficulty with having "none" and "all" categories, with only one intermediate category. To overcome this, a fourth level was included for all attributes to increase the number intermediate levels. Figure 1 shows an example scenario as it appeared in the valuation exercise. The full set of terminology can be found in table 2.

## **Quantitative work: valuation**

Survey data were collected between October 2005 and January 2006. 478 individuals were sampled in total, producing 315 full interviews, in that

respondents reached the end of the questionnaire although not necessarily completing all questions (response rate 66%). Six interviews were abandoned during the interview (1%). Non-responders included refusals to take part (n=105, 22%), unavailable through illness or incapacitation, death or house move (n=37, 8%) and those with whom interviewers were unable to make contact or where the wrong person was interviewed by mistake (n=14, 3%). Of the 315 respondents for whom fully productive interviews were obtained, 255 (53% of the total sample) provided complete best-worst data and are included in the analysis presented here. Characteristics of respondents completing the interview and those who provided complete best-worst data are shown in table 3.

Initial analysis showed that having no control was associated with the lowest attribute level value and so the analyses were initially scaled such that control level one had a value of zero. The impact for each attribute, defined as the average quality of life across all four levels was obtained. All attributes deemed important in the qualitative research phase (Grewal et al., 2006), were also found to be important in this quantitative analysis. The attachment attribute has greatest impact with a utility of 3.23, whilst security has least impact (2.34). Role, enjoyment and control have attribute impacts of 2.87, 2.67 and 2.61 respectively.

Figure 2 shows the results in a different way, in terms of the quality of life for each level of each attribute based on the initial scaling. The values associated with particular levels of each attribute indicate the differences in quality of life that have most value to respondents. For example, there are large differences in value between “none” and “a little”, and between “a little” and “a lot” in terms of love and friendship (attachment). The largest differences in value are not consistent across the levels of particular attributes, with that for security lying between the third and fourth levels (next best to best level) and those for role, control and enjoyment lying between the worst and next to worst levels;

attachment exhibits a large difference between the second (“a little”) and third (“a lot”) levels. For some attributes, particularly enjoyment and role, there is little difference between the third (“a lot”) and fourth (“all”) levels, suggesting only limited additional value for this group of older people for moves between these levels.

Values were then rescaled such that the total value for 11111 sums to zero and the total value for 44444 sums to one. Figures for each attribute level are given in table 2. This rescaling leads to some individual capability levels having negative values (control and attachment). This rescaling does not imply, however, that it is possible for the total capability value to have a value lower than zero as the lowest possible level is 11111 which has a value of zero.

Summing the quality of life scores for each attribute enables estimation of a total score for any given state. Under this rescaling, the value for the capability state which equates to no capabilities, state 11111, has value zero, the state of having a little capability on all attributes (22222) has value 0.556, the state of having a lot of capability on all attributes (33333) has a value of 0.866, and (by definition) having full capability on all attributes (44444) has a value of 1.

Table 4 shows summary statistics for the ICECAP values for all those who provided data on their own capabilities, and for all those who provided data on their own capabilities and also provided complete data for the best-worst scaling exercise.

## DISCUSSION

The data and analysis reported in this paper represent an important step towards developing a capability index for older people for use in economic evaluations that cross the health and social care boundary. The index measure includes attributes that are influenced by health, but also by other factors such as bereavement and finances (other negative influences on quality of life noted in the initial conceptual work (Grewal et al., 2006)), as well as non-health interventions such as the provision of social care.

Not only does the paper focus on the first application of the capabilities approach to the development and valuation of an instrument for use in the economic evaluation of health and social care interventions, it also reports the innovative use of both qualitative and quantitative methods for the development and valuation of this index. The use of qualitative methods to develop terminology for the measure is an unusual step in the development of measures for economic evaluation, but should ensure a greater degree of validity for the measure, in that there is a greater likelihood that it actually measures what it purports to measure. Best-worst scaling was successfully applied to the valuation of quality of life/capability values, with the vast majority of older people who were willing to take part in the survey being able to fully complete the valuation exercise on an interviewer-administered basis. This suggests that this method of valuation may be well-suited to groups who have often found other methods of valuation difficult. Analysis at the population level has been provided here for use in the capability index tariffs. More detailed exploration of values will be possible in future analyses that will consider differences in values by different groups of the population and will conduct analysis at an individual level.

The index draws upon the capabilities literature in some aspects, but not in others. The terminology of the measure fits with the capabilities approach in that individuals are asked about their ability to achieve particular functionings, rather than directly about their functioning. The index measure also draws on the philosophical approach of the capabilities literature in its method of anchoring the measure at no capability and full capability, respectively, which contrasts with the preference elicitation techniques used to value against death in the health economics literature. However, other aspects of the index measure are more strongly influenced by the health economics and health services research literature. First, the measure developed here focuses on individuals' own perceptions of their capabilities, rather than providing some notion of an objective assessment of capability. Whilst this may be feasible for relatively simple and externally imposed capabilities like nutrition levels, the approach of asking others to make an assessment of one's capability for complex capabilities that include an internal component, such as attachment or security, seems unsound. Nevertheless, it may be that "pure" capabilities researchers would refer to this instrument as an index of "perceived capability". This may particularly be the case given the phrasing of two attributes within the index (attachment and enjoyment) in terms of being able to have all the love and friendship/enjoyment and pleasure "that I want". Second, the use of experts as in the capabilities literature (both to develop lists of capabilities and potentially to be involved in their valuation) is an approach that has not been followed. Instead, older people themselves were the sources of attributes and, so far, the source of values. However, there is still the possibility of drawing further upon the capabilities literature in the light of Cookson's suggestion to consider values elicited in the manner described here as a first stage in the development of a final set of values (Cookson, 2005).

There are plans to extend the research on the ICECAP index to use deliberation and debate among a variety of groups to adjust the final set of values used in



the index. This raises issues about how this deliberation and debate should be conducted. For example, who should be included? Should young people's views be sought because of the potential problem of response shift (Verkerk, Busschbach & Karssing, 2001), where expectations among older people may be lowered following adaptation to poorer health states or to other negative influences such as bereavement or lack of income associated with retirement? Potentially the views of a wide variety of groups could be included, such as the general population, older people, charities working with older people, decision makers and so on. Methods could include the use of focus groups to discuss values both from the perspective of an initial valuation and in light of the values obtained here from older people themselves. Reasons for divergence from the set of values obtained from older people directly could be detailed such that the final set could be justified.

In addition to the choices that have been made in this research about the extent to which the approaches discussed within the capabilities literature should be adhered to, there have been other challenges, particularly in relation to terminology. This issue has not been faced previously by capabilities researchers as they have not intended their lists of capabilities to be used by the "final consumers". The index measure proposed here, however, asks individuals to assess their own capabilities in these complex areas where it would be difficult for others to assess them. Although extensive qualitative work around meaning resulted in terminology that is comprehensible to individuals and although people appear to be able to cope with the terminology of capability, there are issues around the meaning of levels that have not been dealt with. The levels settled on here were "all" and "none" (with fairly clear meanings to most people) and, between these extreme levels, "a little" and "a lot". There are two potential problems with using terms such as "a little" and "a lot". One is that these terms are relatively imprecise (as with other similar terminology used in QALYs such as mild, moderate and severe) and "a little" to

one person when measured on a 0 to 100 scale, say, may not be the same as “a little” to another person. This is a common problem in developing measures of this type, although the best-worst scaling approach applied in the main survey allows the assessment of interpersonal differences in values associated with “a little” or “a lot” to be estimated. A second problem is that individuals’ judgements about how they themselves define “a little” or “a lot” on any particular attribute may plausibly be affected by how important these attributes are to that individual. For example, an individual for whom control is very important may require a much greater sense of control to consider themselves to have a lot of control than an individual for whom attachment is much more important. Again this problem will affect almost all such measures.

There are a number of specific avenues for further research. The validity, reliability, sensitivity to change and feasibility of use of the ICECAP index need to be assessed. With the data obtained, there is also the possibility of exploring individual level valuations, an area of research that is ongoing. There are also opportunities to explore the use of deliberative methods to alter values as well as a potential to widen the basis for the instrument to be applicable to the entire adult population rather than just older people. More generally, although this work has addressed the understanding of informants with regard to the use of language indicating that the attribute levels were capabilities, it did not seek to compare terms associated with functionings with those associated with capabilities to determine whether, in fact, informants treat the meanings of these as different or the same. Such a study would potentially be a fruitful area of future research. Finally, this measure considers only capability; death is seen as being no more than the absence of capability (as also in the case of the QALY where death is seen as being no more than the absence of utility). There may be other aspects of death (and particularly of the dying process) that it is not possible to encapsulate within the evaluative space of capabilities and there is clear potential to explore these issues within some other metric.

Notwithstanding these issues, this work has developed an index measure that, potentially, can be used to evaluate health and social care interventions. One option would be to use such an index essentially as a QALY replacement – as a means of retaining a single measure of outcome for such evaluations, but with that measure covering more than just health. Clearly this would currently be feasible only within the context of interventions for older people. A second option for using the index would be as one of a number of outcome measures along the lines of a cost-consequences approach (Coast, 2004). This might well accord better with both a non-welfarist approach to economic evaluation (Coast, 2004) as well as the capabilities approach. As Alkire states “Sen argues that no one principle – for example, efficiency maximisation – suffices for normative economic problems. Rather, a plurality, not only of informational ingredients, but also of combining principles, should be considered” (Alkire, 2005).

Not only does the ICECAP capabilities index for older people draw on a different theoretical approach to evaluation, it has successfully used rigorous methods of qualitative research in the development of its terminology and has innovatively applied the best-worst scaling approach to the valuation of quality of life capability data in a general population. The index provides a new and rigorous approach to considering the value of different interventions for the purpose of economic evaluation.

## **ACKNOWLEDGEMENTS**

This work would not have been possible without the time given up by informants during the qualitative work and respondents to the surveys. Thanks are also due to Ini Grewal and the survey interviewers. This work was supported by the MRC Health Services Research Collaboration.

*Table 1. Socio-demographic information for informants involved in the qualitative work*

	Informants (n=19)
Sex <ul style="list-style-type: none"> <li>• male</li> <li>• female</li> </ul>	7 12
Age	
65-74	12
75-84	6
85+	1
Health status	
No difficulty	5
Physical difficulties	11
Physical and emotional difficulties	3
Living arrangements	
Alone	10
With others	9
Previous occupation*	
Professional/managerial	6
Intermediate	6
Manual	7

\* Classified using occupational groups (Market Research Society, 1991):  
professional/management - A, B; intermediate - C1, C2; manual - D, E.

Table 2: Terminology for attribute levels, and rescaled values, such that the absence of capability, state 1111, is equal to zero, and full capability, state 44444, is equal to one (n=255)

Attribute	Value
<b>Attachment</b>	
I can have all of the love and friendship that I want	0.2535
I can have a lot of the love and friendship that I want	0.2325
I can have a little of the love and friendship that I want	0.1340
I cannot have any of the love and friendship that I want	-0.0128
<b>Security</b>	
I can think about the future without any concern	0.1788
I can think about the future with only a little concern	0.1071
I can only think about the future with some concern	0.0661
I can only think about the future with a lot of concern	0.0321
<b>Role</b>	
I am able to do all of the things that make me feel valued	0.1923
I am able to do many of the things that make me feel valued	0.1793
I am able to do a few of the things that make me feel valued	0.1296
I am unable to do any of the things that make me feel valued	0.0151
<b>Enjoyment</b>	
I can have all of the enjoyment and pleasure that I want	0.1660
I can have a lot of the enjoyment and pleasure that I want	0.1643
I can have a little of the enjoyment and pleasure that I want	0.1185
I cannot have any of the enjoyment and pleasure that I want	0.0168
<b>Control</b>	
I am able to be completely independent	0.2094
I am able to be independent in many things	0.1848
I am able to be independent in a few things	0.1076
I am unable to be at all independent	-0.0512

*Table 3: Characteristics of respondents to the survey.*

	<b>All responders (n=315)</b>	<b>Responders with complete BWS data (n=255)</b>
Number of females (%)	177 (56.2%)	138 (54.1%)
Mean age (SD)	74.6 (6.43)	74.1 (6.01)
Age range (IQ range)	65-95 (70-78)	65-93 (70-78)
Mean EQ-5D score (SD)	0.76 (0.270)	0.77 (0.260)
Number living with spouse (%)	168 (53.3%)	148 (58.0%)
Social class		
Professional (%)	16 (5.2%)	13 (5.2%)
Managerial technical (%)	87 (28.2%)	76 (30.2%)
Skilled non-manual (%)	64 (20.7%)	50 (19.8%)
Skilled manual (%)	63 (20.4%)	50 (19.8%)
Semi-skilled manual (%)	66 (21.4%)	57 (22.6%)
Unskilled manual (%)	13 (4.2%)	6 (2.4%)

*Table 4: Descriptive statistics for ICECAP tariffs among respondents who answered all five capability questions\**

	Full sample (n=313)	Complete BW data (n=253)
Mean	0.814	0.822
Standard deviation	0.152	0.136
Median	0.851	0.852
Inter-quartile range	(0.758,0.914)	(0.769, 0.914)
Maximum	1	1
Minimum	0	0.148
95% confidence interval	(0.797,0.831)	(0.805,0.839)
99% confidence interval	(0.791,0.836)	(0.800,0.844)

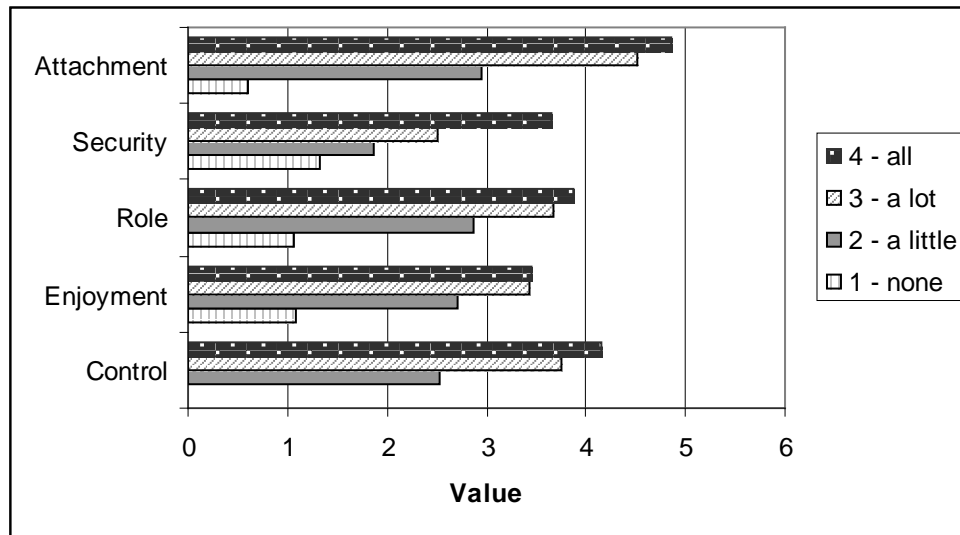
\* Two respondents who completed all choice tasks did not complete all five ICECAP questions



*Figure 1: example scenario*

Quality of life
You can have a lot of the love and friendship that you want
You can think about the future with only a little concern
You are unable to do any of the things that make you feel valued
You can have a little of the enjoyment and pleasure that you want
You are able to be completely independent

Figure 2: Values for the levels of the quality of life attributes obtained from paired random utility multinomial logistic regression, scaled such that control level 1 is equal to zero ( $n=255$ )



## REFERENCES

- Alkire,S. (2005). Why the capability approach? *Journal of Human Development* 6 115-133.
- Anand,P., & van Hees,M. (2006). Capabilities and achievements: an empirical study. *Journal of Socio-Economics* 35 268-284.
- Anand,S., & Sen,A.K. Human Development Index: methodology and measurement. 1994. New York, Human Development Report Office.
- Boreham,R., Airey,C., Erens,B., & Tobin,R. (2002). *National surveys of NHS patients. General Practice 2002*. London: Department of Health.
- Carr-Hill,R.A. (1989). Assumptions of the QALY procedure. *Social Science and Medicine* 469-477.
- Coast,J. (2004). Is economic evaluation in touch with society's health values? *British Medical Journal* 329 1233-1236.
- Cookson,R. (2005). QALYs and the capability approach. *Health Economics* 14 817-829.
- Culyer,A.J. (1990). Commodities, characteristics of commodities, characteristics of people, utilities, and the quality of life. In S.Baldwin (Ed.), *Quality of life. Perspectives and policies* (pp. 9-27). London: Routledge.
- Culyer,A.J. (1989). The normative economics of health care finance and provision. *Oxford.Rev.Econ.Pol.* 5(1), 34-56.
- Department of Health (2001). *National Service Framework for Older People*. London: Department of Health.
- Department of Health, & OPD(PIP) (2004). *Better health in old age*. London: Department of Health.
- Flynn,T.N., Louviere,J., Peters,T., & Coast,J. (2007). Best-worst scaling. What it can do for health care research and how to do it. *Journal of Health Economics* 26 171-189.
- Glaser,B.G., & Strauss,A.L. (1968). *The discovery of grounded theory: strategies for qualitative research*. London: Weidenfeld and Nicolson.

- Glendinning, C. (2003). Breaking down barriers: integrating health and social care services for older people in England. *Health Policy* 65 139-151.
- Grewal, I., Lewis, J., Flynn, T.N., Brown, J., Bond, J., & Coast, J. (2006). Developing attributes for a generic quality of life measure for older people: preferences or capabilities? *Social Science and Medicine* 62 1891-1901.
- Keeney, R.L., & Raiffa, H. (1976). *Decisions with multiple objectives*. New York: Wiley.
- Market Research Society (1991). *Occupation groupings: a job dictionary*. London: Market Research Society.
- Marley, A.A.J., Flynn, T.N., & Louviere, J.J. (2008). Probabilistic models of set-dependent and attribute-level best-worst choice. (In press.). *Journal of Mathematical Psychology*.
- Marley, A.A.J., & Louviere, J.J. (2005). Some probabilistic models of best, worst and best-worst choices. *Journal of Mathematical Psychology* 49 464-480.
- McIntosh, E., & Louviere, J. Separating weight and scale value: an exploration of best-attribute scaling in health economics. Health Economists Study Group, Brunel University. 2002.
- National Institute for Clinical Excellence. Guide to the methods of technology appraisal. 2004. London, National Institute for Clinical Excellence.
- Nussbaum, M.C. (2003). Capabilities as fundamental entitlements: Sen and social justice. *Feminist Economics* 9(2-3), 33-59.
- Robeyns, I. (2005). The capability approach: a theoretical survey. *Journal of Human Development* 6 93-114.
- Robeyns, I. (2003). Sen's capability approach and gender inequality: selecting relevant capabilities. *Feminist Economics* 9(2-3), 61-92.
- Sen, A. (1993). Capability and well-being. In M.C.Nussbaum (Ed.), *The quality of life* Oxford: Clarendon Press.
- Sen, A. (1992). *Inequality reexamined*. New York: Russell Sage Foundation.
- Sen, A. (2005). Human rights and capabilities. *Journal of Human Development* 6 151-166.
- Sen, A. (2002). *Rationality and freedom*. Cambridge, MA: Harvard University Press.

Sen,A. (1982). *Choice, welfare and measurement*. Cambridge, MA: Harvard University Press.

Stata Corporation. Stata Statistical Software. (8.0). 2003. College Station, TX.

Strauss,A., & Corbin,J. (1990). *Basics of qualitative research. Grounded theory procedures and techniques*. London: Sage.

Verkerk,M.A., Busschbach,J.J.V., & Karssing,E.D. (2001). Health-related quality of life research and the capability approach of Amartya Sen. *Quality of Life Research* 10 49-55.

Williams,A. (1985). Economics of coronary artery bypass grafting. *British Medical Journal* 291 326-329.