USING THE HEALTH ASSESSMENT QUESTIONNAIRE-DISABILITY INDEX (HAQ-DI) TO PREDICT THE ICECAP-O CAPABILITY INDEX FOR ARTHRITIS PATIENTS: IS MAPPING FROM PHYSICAL FUNCTION TO CAPABILITY FEASIBLE?

PAUL MITCHELL, TRACY ROBERTS, PELHAM BARTON AND JOANNA COAST

HEALTH ECONOMICS UNIT, UNIVERSITY OF BIRMINGHAM, B15 2TT, UK.

E-mail Corresponding author at: PXM902@bham.ac.uk

ABSTRACT

AIMS
The capability approach is an alternative theoretical basis to the more traditional welfare economic approach used to assess individual well-being. For health economics, the capability approach offers an alternative basis for health service evaluation including expanding the evaluative space to a broader measurement of quality of life and moving the decision making rules beyond the sole focus on health maximisation [Coast et al. 2008a]. As yet, there are only a few studies that have included one of the ICECAP indices, so one challenge in using the approach in practice is to explore other possibilities for obtaining ICECAP data. Mapping is an inferior technique to primary data collection, but can be useful to predict values for a measure which has not been collected within a study dataset. The mapping process is carried out in a separate estimation dataset, which calculates the statistical relationship between the target measure and the starting measure where both were collected within this separate dataset [Brazier et al. 2010]. This study aims to examine whether mapping techniques used to predict utilities for economic evaluations, can also be used to measure capability well-being.

METHODS
101 patients with moderate to severe arthritis completed questionnaires for the Health Assessment Questionnaire-Disability Index (HAQ-DI – the starting measure) and the ICECAP-O capability index (the target measure). HAQ-DI is a common measure in arthritis trials and is the primary predictor of utilities in rheumatoid arthritis mapping studies [Marra et al. 2011]. Descriptive statistics were used to analyse what regression family would be most appropriate for this dataset. Socio-demographic characteristics (age and gender) were included in regressions to examine their explanatory power for capability. Model specification focused on whether overall scores or dimension scores were best able to capture the relationship between the two measures.

DATA
The mean age of patients providing data was approximately 70 years old (range 48-89). Mean HAQ-DI scores were 1.774, measured on a scale of 0 (best) to 3 (worst). Mean ICECAP-O scores were 0.779, where capability is measured on a 0 (no capability) to 1 (full capability) scale.

RESULTS
Early results suggest that in this dataset HAQ-DI, as the sole explanatory variable of ICECAP-O, is the most appropriate mapping relationship, as it produced the lowest measurement error statistics (Root Mean Squared Error and Mean Absolute Error). Age and gender are not significant explanators of capability. R^2 statistics, used as measures of goodness-to-fit in regression analysis, show that the predictive relationship between HAQ-DI and ICECAP-O is low ($R^2 = 0.248$)

DISCUSSION
While this is work-in-progress, focusing solely on a physical ability measure appears to produce relatively weak predictions of capability scores. Additional explanatory variables for capability should be explored and tested in future mapping exercises involving the ICECAP measures. We are keen to explore the views from HESG and CES members about the potential implications of this research for measuring capability within health economic evaluations.

BACKGROUND

It has been argued within the health economics literature that an evaluation framework based primarily on the capability approach would generate richer data for quality of life comparisons across interventions in the health service (Verkerk et al. 2001; Anand & Dolan 2005; Coast et al. 2008a; Kinghorn & Smith 2008; Ruger 2010). Applying the capability approach to health economics would allow for an assessment of broader benefits of quality of life than are currently measured in health economic evaluations presently, through health related quality of life (HRQOL) questionnaires such as the EQ-5D (Brooks et al. 1996), HUI (Furlong et al. 2001) and the SF-6D (Brazier et al. 2002).

The ability to provide a consistent theoretical basis in health economic evaluations is not simple using current methodology, given the various influences of welfare economics, capability theory and subjective well-being to produce the QALY outcome measure from such evaluations (Cookson & Culyer 2010). This has led to heterogeneous methods used to generate QALYs which are by no means comparable, a variety of techniques exerted to elicit values or preferences by trading between quality and quantity of life and also an inability to account for the distribution of healthcare in a society (Nord et al. 2009).

The capability approach developed predominantly by Amartya Sen (1985, 1992, 1993) offers an alternative theoretical basis for societal welfare assessment than the more traditional utilitarian approach used in welfare economics. Instead of focusing on what a person does and how they function, the capability approach expands the evaluative space and asks what a person can do and what they are capable of doing.

For an example of the difference between focusing on a person's capabilities rather than their functionings, consider two individuals (A & B) both staying in their homes for the evening (example from Sen 2009). Whilst individual A has the choice to leave their home for the evening, individual B has no choice but to stay at home as they are under house arrest. If we calculate both individuals' well-being in terms of their functionings from staying at home, both individuals have equivalent outcomes. However, individual A is in a better position than individual B as they have the capability to choose to go out and socialise for the evening (Cookson 2005, has another good example of equal functionings but varying capability sets when he compares voting rights of politically apathetic individuals in a democracy to those who are unable to vote in an autocratic state). By assessing individual capability sets as opposed to achieved functionings solely, we can distinguish between the differences of

well-being in individuals A and B, which we would not be able to do if we focused solely on their functionings.

While the case for using the capability approach is strong theoretically, the direct empirical application of the approach has been quite peripheral, particularly within the health economics discipline (Lorgelly et al. 2010). The ICECAP capability indices (ICECAP-O by Coast et al. 2008c and the ICECAP-A by Al-Janabi et al. 2011a) offer means for measuring capability well-being in health economics. Both the ICECAP-O (designed for the over 65s) and the ICECAP-A (designed for the adult population over 18) are simple self-report questionnaires each containing five questions with four levels asking about individual well-being in terms of capability. Nonetheless, unlike the EQ-5D, the ICECAP indices move away from focusing solely on health related quality of life. Instead the capability measures ask questions related to broader considerations of quality of life.

For the ICECAP-O, the five attributes of capability well-being assessed are attachment, security, role, enjoyment and control, while for the ICECAP-A the categories are stability, attachment, autonomy, achievement and enjoyment (see Appendix I for sample ICECAP-O questionnaire). Both questionnaires were developed using qualitative methods with samples of the UK population (ICECAP-O: Grewal et al. 2006; ICECAP-A: Al-Janabi et al. 2011a). The construct validity of the ICECAP-O has been tested (Coast et al. 2008b; Flynn et al. 2011), as has the conceptual validity of the ICECAP-A measure compared with the EQ-5D (Al-Janabi et al. 2011b).

The assessment of capability well-being would preferably be informed by primary data, given that such research is novel within health economics and analysis at this time will be mainly exploratory. The ICECAP indices are in their early stages of application, so the absence of primary data for measures such as the ICECAP-O and ICECAP-A in trials and interventions makes the task of probing the use of the capability approach more complex.

An inferior technique to primary data collection called mapping or cross-walking has been used to predict HRQOL scores when primary data have been unattainable. Mapping allows for the prediction of one measure (the target measure) by defining a statistical relationship with another measure (the starting measure). This starting measure must be collected in both an estimation and study dataset, so that the target measure, which was collected only in the estimation dataset, can be used to derive projections of the target measure in the study dataset (Brazier et al. 2010).

This paper explores the applicability of using the mapping process to infer statistical values of capability well-being. This will be carried out by mapping onto the ICECAP-O capability index from a generic measure of health functioning, the health assessment questionnaire-disability index (HAQ-DI). The HAQ-DI is a common measure of reduced functionality of patients with rheumatoid arthritis and has been used in numerous mapping studies using HRQOL questionnaires as the target measure (Marra et al. 2011).

The aim of this research is to assess the feasibility of mapping from a functioning-based health measure to a capability-based generic outcome measure. This paper provides evidence as to whether such a process could be advocated in, for example, economic evaluations based on secondary data and decision modelling.

METHODS

OUTCOME MEASURES

Starting Measure (HAQ-DI)

The health assessment questionnaire (HAQ) is primarily a measurement of functional disability in patients with Rheumatoid Arthritis. The HAQ has been frequently used, refined and modified since its inception over thirty years ago (Fries et al. 1980). While the "full HAQ" incorporates five dimensions of outcome (Known as the five Ds: "disability", "discomfort", "drug side effects", "dollar costs" and "death"), the most widely used aspect of the HAQ is the disability index, the HAQ-DI (Bruce & Fries 2003), which is the primary interest in this study. In the traditional HAQ-DI format, there are 20 questions covering 8 areas of functioning, with 2 to 3 questions for each functioning (see Appendix II). The highest (worst) score for each area of functioning is used to calculate the HAQ-DI score, with each of the eight categories valued equally. Additionally, polar questions regarding the extra support patients require from individuals, aids and devices are sometimes also taken into consideration. When extra assistance is required in a HAQ-DI category, a score of 2 (much difficulty) is recorded for that category, unless 3 (unable to do) has already been reported. The final HAQ-DI score for each person ranges from 0 (no problems functioning) to 3 (not able to function). A simpler measure of the HAQ-DI called the modified HAQ or MHAQ, which only asks one question per functioning (8 in total) has been developed more recently and is now used interchangeably with the original HAQ-DI (Wolfe 2001).

Target Measure (ICECAP-O)

The ICECAP-O capability index is a simple, self-report five-part questionnaire designed to measure quality of life in terms of capability well-being for older people (Appendix I). The measure was developed primarily for use in resource allocation decisions in health care for the over 65s, although there is no reason why the measure could not be used in other disciplines assessing quality of life for the elderly (Coast et al. 2008c).

The ICECAP-O consists of five attributes of capability well-being measured across four levels, which were developed through an iterative process using qualitative research methods (Grewal et al. 2006). The ICECAP-O measures capability well-being in terms of:

➢ Attachment   –   Love & Friendship
➢ Security   –   Thinking about the future without concern
➢ Role   –   Doing things that make you feel valued
➢ Enjoyment   –   Enjoyment and Pleasure
➢ Control   –   Independence

Capability is anchored to a 1-0 scale, where 1 is equivalent to full capability on all attributes (44444) and 0 to having no capability on any of the five quality of life dimensions (11111). Obtaining values for each of the five capability well-being attributes was carried out in a separate valuation study (Coast et al. 2008c). The construct validity of the measure has been established (Coast et al. 2008b; Flynn et al. 2011), while the first applied study using the measure has recently been published (Makai et al. 2011).

DATASET SELECTION

Estimation Dataset:

In order to predict statistical association between two measures, the target measure (ICECAP-O) and starting measure (HAQ-DI) were required to be collected within the same estimation dataset. Given the paucity of studies using the ICECAP indices so far, the first step in selecting an appropriate dataset required ICECAP data to have been collected. Two studies using ICECAP-O and one study using the newer ICECAP-A were identified, which collected scores of capability wellbeing for arthritis patients. However, none of the three studies collected HAQ-DI in its original format. Only one study, consisting of

mainly osteoarthritis but some rheumatoid arthritis patients who required hip and knee replacements, collected physical function data, primarily from the WOMAC questionnaire (Bellamy et al. 1988). However, additional questions were also collected from the HAQ-DI and the generic SF-36 (Ware & Sherbourne 1992), but were restructured using the same phraseology as the WOMAC questionnaire (Pollard et al. 2009). This dataset was chosen for further analysis.

Adjusted HAQ-DI (see Appendix III)

HAQ-DI scores were processed through responses given in the physical function questions section collected alongside the ICECAP-O. A high correlation between HAQ-DI, WOMAC and physical function questions on the SF-36 have been found previously (see Bruce & Fries 2003 for examples).
Out of the twenty HAQ-DI questions from the eight areas of physical functioning:

-         11 physical function questions were worded in a directly comparable manner
-         2   questions required 2 physical function questions (i.e. 4 questions for 2 HAQ-DI questions)
-         7   questions and 2 categories (eating & grip) had no comparators in this dataset

Given that the eating and grip categories had no direct matches in our dataset, substitute questions were considered for both categories. However, such substitution introduced the potential for double-counting some of the physical functioning categories in the HAQ-DI. Since only six out of eight categories are required to record a HAQ-DI score (Bruce & Fries 2003), no substitute questions were used and we proceeded with fifteen questions for six categories.

Study Dataset:

The study dataset is where the mapping process carried out in the estimation dataset is applied to a case study where the target measure has not been collected prospectively. Although this application is not reported here, the most recent version of the Birmingham Rheumatoid Arthritis Model (BRAM), an individual sampling model which processes different drug treatment strategies available for patients (Malottki et al. 2011), will be used for further analysis as our study dataset. The BRAM model adopted mapping techniques to predict EQ-5D scores from the HAQ-DI in order to generate QALYs (Barton et al. 2004).

MAPPING

Mapping or cross walking is a technique used to predict scores for the target measure (ICECAP-O) through estimating the statistical relationship between the target measure and the starting measure (HAQ-DI). This estimation dataset, where the relationship between the two measures is derived through statistical association, allows for predictions of the target measure to be mapped into the study dataset, where the starting measure was also collected (Brazier et al. 2010). Thus, the mapping process permits the estimation of values for measures of outcome used in economic evaluation where no direct estimates exist.

STATISTICAL ANALYSIS

Statistical analysis followed similar methods to those used in past mapping studies, in particular Kaambwa (2009) and Kaambwa et al. (2011).

The first step in the statistical analysis involved the generation of descriptive statistics for all possible dependent and independent variables. Explanatory variables used in the mapping process were required to have been collected in the study dataset (i.e. the BRAM model), as mapping from other variables could not be used within the study dataset. There were three explanatory variables considered: overall HAQ-DI scores, age and gender. Potential predictor variables were overall ICECAP-O scores, treated as continuous data or the five dimensions of ICECAP-O, treated as categorical data. Scattergraphs were used to explore the ICECAP-O scores with each potential explanatory variable and box-plots for each of the five attributes of capability well-being were used to ascertain the relationship between each attribute level and overall HAQ-DI scores.

Since there was no a priori knowledge of which regression model would best predict ICECAP-O scores, the second step involved identifying all possible regression models which could be included into the estimation dataset. While six different types of model specifications were identified by Brazier et al. (2010) (reproduced here as Table 1 for illustrative purposes), only two model specifications (model specification 1 and 6) were considered in our analysis, due to the restrictive nature of the data collected in the estimation dataset but also the summary nature of the data contained in the study dataset.

ICECAP-O as continuous dependent variable

OLS regressions were considered for the first model specification from Table 1, as they were used in predicting EQ-5D scores for the BRAM model previously (Barton et al. 2004). There are limitations using this regression model family when ceiling effects are present in the data, leading to biased estimates (Gray et al. 2006). However, there were no such ceiling effects present in the ICECAP-O data. Given that most mapping studies have used OLS regressions (Brazier et al. 2010), we proceeded with this family for our study.

For model specification 1 in Table 1, overall ICECAP-O scores were treated as a continuous dependent variable for all regressions. Stepwise regression is a statistical process of finding the best model fit for a regression, given no *a priori* knowledge of what explanatory variables should and shouldn't be included, thus limiting the chance of model over-specification and omitted variable bias (Draper & Smith 1981). In the form used here, stepwise regression starts with all possible explanatory variables in the original regression, before removing the least significant explanators for each model re-run, until only statistically significant predictors of the dependent variable are left (Cohen 2003). Once all irrelevant variable(s) are eliminated from the regression model, tests are carried out on the residuals of the explanatory variables to test for normality, linearity and heteroscedasticity within the dataset. If such problems are contained within datasets, variable transformation is required to produce BLUE (Best Linear Unbiased Efficient) estimates (Gujarati 2004).

ICECAP-O attributes (5) as categorical dependent variables

Using model specification 6 on Table 1, where the five ICECAP-O categorical dimensions are treated as dependent variables, there are two options for further analysis: ordinal or multinomial logistic regression. Ordinal logistic (OL) regression has the benefit of recognizing that the categorical data are ordered. However, it requires the assumption of proportional odds between categories, so that all levels within a category have parallel slopes (Agresti 2002).

Alternatively, if the proportional odds assumption is violated, multinomial logistic (ML) regression can be used instead. However, this type of regression does not recognise the order of categories in the same way as OL regressions. Instead, probabilities are assigned to each level within a category, so that for a given HAQ-DI score, the chances of having full, a lot, a little or no capability (the four levels for each capability attribute – see Appendix I) are predicted.

Three types of model prediction using ML regression are explored:

- Direct: using the probability of every level for each category level to predict overall score (Tsuchiya et al. 2002)

- Indirect: using the highest probability level in each category to predict overall score (Tsuchiya et al. 2002)

- Response: using Monte Carlo simulations to generate predicted ICECAP-O responses (Gray et al. 2006)

A fourth method has recently been applied known as "probabilistic mapping" using Bayesian networks (Le & Doctor, 2011). However, since probabilistic mapping relies on categorical data predicting categorical data and the study dataset relies on overall HAQ-DI scores, rather than HAQ-DI categories, this method was not pursued any further in this research. Future mapping studies should include this procedure if enough data are available to proceed with this method.

Measuring predictive ability of regression models

The most common measure of goodness-to-fit of a regression model is the R^2 statistic, which measures the explanatory power of the independent variables on the dependent variables (Gujarati 2003). However, in mapping studies, the primary concern is not on the explanatory power of regression models but in the ability to predict the dependent variable from the model. We use two measures of predictive error in this study:

- Mean absolute error (MAE). The absolute difference between the predicted and the observed value (Greene 2002).

- Root mean squared error (RMSE). The square root of the difference between the predicted and observed value squared (Greene 2002).

All models were tested for model specification error. Normality and heteroscedasticity tests were also carried out. STATA version 9 (StataCorp, 2006) and Microsoft Excel (2003) were used for predicting models.

RESULTS

Table 2 shows the demographics of the population under consideration. Main findings from this table show that the mean ICECAP-O score for the 101 people was 0.779, which is less than the mean of the population used in the valuation study of ICECAP-O of 0.814 (Coast et al. 2008). The mean HAQ-DI score was 1.773, with an average age of approximately 70 years. Figure 1 shows the relationship between ICECAP-O scores and HAQ-DI scores in the population. From Figure 1, it is clear that there is skewness in the ICECAP-O data towards higher ICECAP-O scores. This signalled from an early stage in analysis that variable transformation would be required for this data. Figures 3 to 7 are box-plots of the five attributes of ICECAP-O, looking at the variation in HAQ-DI scores for the four levels on each ICECAP-O attribute. Exact numbers of the ICECAP-O level recorded from our population sample on each of the five attributes of capability in this sample can be found in Table 3.

When examining p-values in Table 5, the relationship between full capability and no capability as the base level, HAQ-DI is significantly different for four of the five ICECAP-O attributes (excluding the "attachment" attribute), while the enjoyment attribute was the only attribute able to capture a significant difference between no and a little capability (see also Figures 3-7 for a visual analysis of the data). The inability to capture significant differences for some ICECAP-O levels can be partially explained by the low numbers recording a "no capability" score for any of the five capability well-being attributes (see Table 3). The ability to predict any of the attachment levels from the HAQ-DI does not seem possible from these findings.

In total, six models predicted ICECAP-O scores or dimensions. Three models predicted overall ICECAP-O scores as a continuous variable through Ordinary Least Squares (OLS) regression (Models 1-3 on Table 4 & Table 6). Model 1 considered HAQ-DI score, age and gender as explanatory variables of overall ICECAP-O scores. The application of stepwise regression left HAQ-DI as the only significant explanatory variable of ICECAP-O. This relationship between HAQ-DI and ICECAP-O is Model 2. When testing Model 2 for normality, linearity and heteroscedasticity, the residuals were found to be non-normally distributed and heteroscedastic. Data transformation is used to address such problems which produce inefficient estimates. The cubic function of ICECAP-O was found to address the problems of non-normality (which can be seen in Figure 2) and heteroscedasticity. The relationship between ICECAP-O^3 and HAQ-DI is predicted in Model 3.

Three regression models explored ICECAP-O dimensions as dependent variables. The assumption of proportional odds between categorical levels did not hold for the ICECAP-O dimensions using the Wald test, so ordinal logistic regression was not used here. The three methods of multinomial logistic (ML) regression (direct - Model 4, indirect - Model 5 and response - Model 6 in Table 6) were analysed for their ability to predict ICECAP-O categories with HAQ-DI as the sole explanatory variable (Table 5)

Table 6 presents predictive statistics for our six regression models. Model 1 recorded the lowest mean ICECAP-O estimate of 0.776, while Model 5 (indirect) recorded the furthest estimate from the observed mean at 0.826. Two models (Model 3 and Model 6-response) were able to predict the highest observed capability score (i.e. 44444 or ICECAP score of 1), while Model 4 (direct) was most distant from this value with its highest score being 0.839. The two Models (3&6) which produced the highest estimates also predicted the lowest ICECAP-O scores (0.251 and 0.000 respectively), while the other four models' minimum scores were all greater than 0.55. Given the limited range of these four models, Model 3 and Model 6 were the only models considered for further analysis in terms of prediction error.

While the ranges of both Models 3 and 6 are comparable with the observed ICECAP-O scores, the predictive ability of the models is not comparable, with the response mapping (Model 6) recording lower scores for both the root mean squared error (RMSE) and mean absolute error (MAE). Indeed, the response mapping model outperforms all other models examined here in terms of RMSE and MAE (see Table 6). Given our primary concern is the predictive ability from our regressions, this analysis suggests the response mapping approach (Model 6) outperforms the five other model types examined here.

DISCUSSION

In this study, we have investigated the ability of econometric regression models to predict capability well-being in terms of ICECAP-O from a health functioning questionnaire (HAQ-DI) for arthritis patients. Here, the multinomial logistic response mapping approach produces the lowest predictive errors from the starting measure (HAQ-DI overall score) onto the target measure (ICECAP-O levels). Comparing results with the ICECAP-O construct validity study (Coast et al 2008b), the results in this paper would add strength to the argument of a strong relationship between the role, enjoyment and control attributes on the ICECAP-O index with measures of physical function (Table 5). However, the attachment attribute, like in the construct validity paper, did not show a strong relationship with the HAQ-DI, so this would suggest

that more than measures of physical functioning would be required to predict all attributes on the ICECAP-O index accurately.

There are a number of limitations within this work. Only two out of six model specifications were explored from Table 1 due to lack of relevant data. The HAQ-DI scores were inferred from comparable questions on the WOMAC questionnaire among others (Appendix III). While this is not ideal, it was the best information available to generate HAQ-DI scores. The sample size of 101 is small compared to past mapping studies using the response mapping approach (e.g. Gray et al. 2006), so 100 Monte Carlo simulations of the dataset were required. However, there have been studies with smaller sample sizes in the past (Brazier et al. 2010). Ideally, an equal spread of HAQ-DI scores across the sample would have been useful like previous mapping studies (Barton et al. 2004). For the multinomial logistic (ML) regressions (Models 4-6 on Table 6), it would have been more acceptable to have greater numbers in the lowest levels of capability, as it would have produced more reliable results. In mapping studies, there is usually a separate validation dataset to explore the findings from the prediction dataset, where the statistical relationship between the two measures is explained. No separate dataset was available to validate our mapping results, so the same dataset was used for predicting and validating the regression models.

This research looked at all potential mapping algorithms which have been used in previous studies. We explored the use of both ordinary least squares (OLS) and multinomial logistic (ML) regression, finding that the response mapping approach proposed by Gray et al. (2006) produced the lowest predictive errors for ICECAP-O scores. However, a new approach using Bayesian networks has produced lower predictive errors than the response mapping approach using ML regressions and also for OLS predictions too (Le & Doctor 2011). This Bayesian network approach used levels from a generic health questionnaire (SF-12) to predict levels for a commonly used measure in health economic evaluations, the EQ-5D. Since not all HAQ-DI levels were captured in our overall HAQ-DI score and our study dataset (the BRAM model) only uses overall HAQ-DI scores, this method was not explored in detail here. It is intended to explore this method when the necessary information becomes available to the researcher.

The ICECAP-O capability index offers a broader measure of quality of life for use within economic evaluations than the HRQOL questionnaires used presently. This paper shows how mapping techniques could be used to explore the application of the capability approach within health economics. The use of mapping procedures is comparably inexpensive and less time consuming than primary data collection for complex interventions, whose treatment lasts for many years. Therefore, mapping techniques, such as

those explored here, could be used for justifying primary data collection of new questionnaires such as the ICECAP indices in future prospective data collection studies.

The application of the capability approach to health economics necessitates the consideration of the distribution of health, alongside efficiency outcomes in an evaluation format (Coast et al. 2008a). Techniques used in human development and public economics research (Alkire & Foster 2011) should be explored to examine which interventions are targeting people who have insufficient capability. The mapping process allow for further exploration of these issues within health economics.

Given the limitations associated with this single study, it would be inappropriate to state definitively whether or not it is feasible to capture capability well-being to a sufficient level from a physical functioning questionnaire. Future research will focus on the application of the mapping work in this paper into the BRAM model to explore how the ICECAP indices can be used to measure both efficiency and distribution health outcomes from complex interventions. In future mapping studies using a measure of physical functioning to predict capability well-being, socio-demographic variables should be explored alongside measures of physical functioning, to examine the variety of influences which are required to predict a broader measure of quality of life in terms of capability well-being such as the ICECAP-O.

References

Agresti A (2002) Categorical Data Analysis (2nd ed.). Wiley Interscience: New Jersey.

Al-Janabi H, Flynn TN, Coast J (2011a) Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Quality of Life Research* DOI: 10.1007/s11136-011-9927-2

Al-Janabi H, Keeley T, Mitchell P, Coast J (2011b) Completion of self-reported health and capability measures: a think aloud study. Health Economists' Study Group, Bangor, UK.

Alkire S & Foster J (2011) Counting and multidimensional poverty measurement. *Journal of Public Economics* 95: 476-487.

Anand P & Dolan P (2005) Equity, capabilities and health. *Social Science & Medicine* 60(2): 219-222

Barton P, Jobanputra P, Wilson J, Byran S, Burls A. (2004) The use of modelling to evaluate new drugs for patients with a chronic condition: the case of anti-bodies against tumour necrosis factor in rheumatoid arthritis. *Health Technology Assessment* 8(11): 1-104.

Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW (1988) Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis in the knee or the hip. *The Journal of Rheumatology* 15(12):1833-40.

Brazier J, Roberts J, Deverill M (2002) The estimation of a preference based measure of health from the SF-36. *Journal of Health Economics* 3: 271-292.

Brazier J, Yang Y, Tsuchiya A, Rowen D (2010) A review of studies mapping (or cross-walking) non-preference based measures of health to generic preference based measures. *The European Journal of Health Economics* 11(2): 215-225

Brooks R. and the EuroQol Group (1996) EuroQol: the current state of play *Health Policy* 37(1): 53-72.

Bruce B & Fries JF (2003) The Stanford Health Assessment Questionnaire: A Review of Its History, Issues, Progress, and Documentation. *Journal of Rheumatology* 30: 167-78.

Coast J, Smith R, Lorgelly P (2008a) Should the capability approach be applied in health economics? *Health Economics* 17(6): 667-670.

Coast J, Peter T, Natarjan L, Sproston K, Flynn T (2008b) An assessment of the construct validity of the descriptive system for the ICECAP capability measure for older people. *Quality of Life Research* 17(7): 967-976.

Coast J, Flynn T, Natarajan L, Sproston K, Lewis J, Louviere J, Peters T (2008c) Valuing the ICECAP capability index for older people *Social Science & Medicine* 67(5): 874-882.

Cohen J (2003) Applied multiple regression/correlation analysis for the behavioural sciences (3rd ed.) London: L. Eribaum Associates.

Cookson R (2005) QALYs and the capability approach. *Health Economics* 14: 817-829.

Cookson R & Culyer A (2010) Measuring overall population health: The use and abuse of QALYs. In: Killoran A. And Kelly M. (eds.) Evidence-based Public Health: effectiveness and efficiency 148-168.

Draper NR & Smith H (1981) Applied Regression Analysis (2nd ed.) Chichester, Wiley.

Flynn TN, Chan P, Coast J, Peters T (2011) Assessing Quality of Life among British Older People Using the ICEPOP Capability (ICECAP-O) Measure. *Applied Health Economics and Health Policy* 9(5): 317-329.

Fries JF, Spitz PW, Kraines RG, Holman HR. (1980) Measurement of patient outcomes in arthritis. *Arthritis Rheumatology* 23:137-145.

Furlong WJ, Feeny DH, Torrance GW, Barr RD (2001) The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Annals of Medicine* 33(5): 375-384.

Gray AM, Rivero-Arias O, Clarke PM (2006) Estimating the Association between SF-12 Responses and EQ-5D Utility Values by Response Mapping. *Medical Decision Making* 26: 18-29.

Greene WH (2002) Econometric Analysis (4th Edition) Upper Saddle River, NJ: Prentice Hall.

Grewal I, Lewis J, Flynn T, Brown J, Bond J, Coast J (2006) Developing attributes for a generic quality of life measure for older people: Preferences or capabilities? *Social Science & Medicine*, 62(8): 1891-1901.

Gujarati D (2004) Basic Econometrics (4th ed.) Boston: McGraw-Hill.

Hochberg MC, Chang RW, Dwosh I, Lindsey S, Pincus T, Wolfe F (1992) The American College of Rheumatology 1991 revised criteria for the classification of global functional status in rheumatoid arthritis. *Arthritis and Rheumatism* 35: 498-502.

Kaambwa BC (2009) Ch.9 Challenges with outcome data: empirical analysis 236-310, In: Statistical issues in service evaluation – a case of intermediate care. Ph.D. thesis, University of Birmingham.

Kaambwa B, Billingham C, Byran S (2011) Mapping utility scores from the Barthel Index. *European Journal of Health Economics* DOI 10.1007/s10198-011-0364-5

Kinghorn P & Smith R (2008) From Theory to Practice: Are we capable of operationalizing the capability approach? Health Economists' Study Group, Norwich, UK.

Le QA, Doctor JN (2011) Probabilistic Mapping of Descriptive Health Status Responses Onto Health State Utilities Using Bayesian Networks. An Empirical Analysis Converting SF-12 Into EQ-5D Utility Index in a National US Sample. *Medical Care* 49(5): 451-460.

Lorgelly PK, Lawson KD, Fenwick EAL, Briggs AH (2010) Outcome Measurement in Economic Evaluations of Public Health Interventions: a Role for the Capability Approach? *International Journal of Environmental Research and Public Health* 7(5): 2274-2289.

Makai P, Brouwer WBF, Koopmanschap MA, Nieboer AP (2011) Capabilities and quality of life in Dutch psycho-geriatric nursing homes: an exploratory study using a proxy version of the ICECAP-O. *Quality of Life Research.* DOI 10.1007/s11136-011-9997-1

Marra CA, Bansback N, Anis AH, Shojania K (2011) Introduction to economic modelling for clinical rheumatologists: application to biologic agents in rheumatoid arthritis. *Clinical Rheumatology,* 30(s1): 9-18.

Pollard B, Dixon D, Dieppe P, Johnston M (2009) Measuring the ICF components of impairment, activity limitation and participation restriction: an item analysis using classical test theory and item response theory. *Health and Quality of Life Outcomes,* 7(1): 41-60.

Ruger JP (2010) Health Capability: Conceptualization and Operationalization. *American Journal of Public Health* 100(1): 41-49.

Sen A (1985) Commodities and Capabilities. Oxford: Oxford University Press.

Sen A (1992) Inequality Reexamined. Oxford: Clarendon Press.

Sen A (1993) Capability and Well-Being. In: Nussbaum M & Sen A. eds. The Quality of Life, 30-53. Clarendon Press: Oxford.

Sen A (2009) The Idea of Justice. London: Allen Lane.

StataCorp (2006) Intercooled stata 9 for windows. Stata Corporation, College Station, TX.

Tsuchiya A, Brazier J, McColl E, Parkin D (2002) Deriving preference-based single indices from non-preference based condition-specific instruments: Converting AQLQ into EQ5D indices. Sheffield Health Economics Group Discussion Paper 02/1. University of Sheffield: ScHARR.

Ware JE, Sherbourne CD (1992) The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection. *Medical Care* 30(6): 473-482.

Wolfe F (2001) Which HAQ is best? A comparison of the HAQ, MHAQ and RA-HAQ, a difficult 8 item HAQ (DHAQ) and a rescored 20 item HAQ (HAQ20): an analysis of 2,491 rheumatoid arthritis patients following leflunomide initiation. *The Journal of Rheumatology* 28(5): 982-989

Acknowledgements

Tables & Figures

Table 1. Model specifications for mapping functions (from Brazier et al. 2010)

Model	Dependent Variable	Discrete (D) or Continuous (C)	Independent Variables Main Effects	Discrete (D) or Continuous (C)	Interactions	Other Measures
1	Index	C	Overall Score	C		For any model:
2	Index	C	Dimension Scores	C	Dimensions	Squared terms,
3	Index	C	Item Levels	C	Items	other health
4	Index	C	Item Levels	D	Items level	measures,
5	Dimension level	C	Models 1-4	C/D	Models 1-4	clinical measures,
6	Dimension level	D	Models 1-4	C/D	Models 1-4	demographics

Table 2. Demographics and key variable characteristics of sample:

n	101	
Age- mean & SD	69.983	(8.847)
Males (n)	54	
Employed (n)	13	
Living Alone (n)	27	
ICECAP-O mean & SD	0.779	(0.162)
HAQ-DI mean & SD	1.774	(0.598)

Table 3. Distribution of observed ICECAP answers across dataset:

Capability	No (1)	A Little (2)	A Lot (3)	Full (4)	Total
attachment	2	9	36	54	101
security	10	28	48	15	101
role	5	37	46	13	101
enjoyment	10	34	46	11	101
control	4	27	44	26	101
% share of capability	6.14	26.73	43.56	23.56	100

Table 4. Ordinary Least Squares (OLS) models

Models	Model 1 (icecap)		Model 2 (icecap)		Model 3 (icecap³)	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
Total HAQ-DI	-0.131	0.024	-0.135	0.024	-0.22	0.035
Age	0.002	0.002				
Female	-0.018	0.029				
Constant	0.882	0.122	1.018	-0.135	0.918	0.653
R-Squared		0.266		0.248		0.286

Table 5. Multinomial Logistic (ML) Models with ICECAP-O levels & overall HAQ-DI score
(Level 1 on ICECAP-O, i.e. no capability, base case for all 5 capability attributes)

	Attachment					
Capability	A Little		A Lot		Full	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
HAQ-DI	-1.681	1.450	-1.792	1.367	-1.542	1.353
Constant	4.934	3.328	6.512	3.204	6.478	3.187
p-values	0.246	0.138	0.190	0.042	0.254	0.042

	Security					
Capability	A Little		A Lot		Full	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
HAQ-DI	-1.109	0.695	-2.081	0.696	-1.822	0.785
Constant	3.352	1.561	5.627	1.531	4.016	1.668
p-values	0.111	0.032	0.003	0.000	0.020	0.016

	Role					
Capability	A Little		A Lot		Full	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
HAQ-DI	-1.449	0.990	-2.658	1.019	-3.482	1.152
Constant	5.199	2.383	7.608	2.407	7.592	2.534
p-values	0.143	0.029	0.009	0.002	0.003	0.003

	Enjoyment					
Capability	A Little		A Lot		Full	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
HAQ-DI	-1.756	0.734	-2.764	0.764	-3.023	0.926
Constant	4.905	1.710	6.964	1.732	5.936	1.922
p-values	0.017	0.004	0.000	0.000	0.001	0.002

	Control					
Capability	A Little		A Lot		Full	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
HAQ-DI	1.011	1.153	-2.233	1.169	-3.549	1.264
Constant	-0.351	2.545	6.615	2.472	8.036	2.566
p-values	0.380	0.890	0.056	0.007	0.005	0.002

Table 6. Predictive error results: OLS (Models 1-3) & ML (Models 4-6) regressions

				Predicted versus observed ICECAP scores			
	Description	Mean	Max	Min	Correlation	RMSE	MAE
Observed		0.779	1.000	0.159	-	-	-
Model 1	see Table 4	0.776	0.946	0.608	0.498	0.140	0.103
Model 2	see Table 4	0.779	0.968	0.665	0.497	0.139	0.101
Model 3	see Table 4	0.808	1.000	0.251	0.530	0.207	0.165
Model 4	Direct mapping	0.768	0.837	0.587	0.503	0.141	0.106
Model 5	Indirect mapping	0.829	0.914	0.574	0.462	0.153	0.101
Model 6	Response mapping	0.779	1.000	0.000	0.509	0.138	0.099

Figure 1. Relationship between ICECAP-O & HAQ-DI: OLS (Model 2 in Table 4&6)
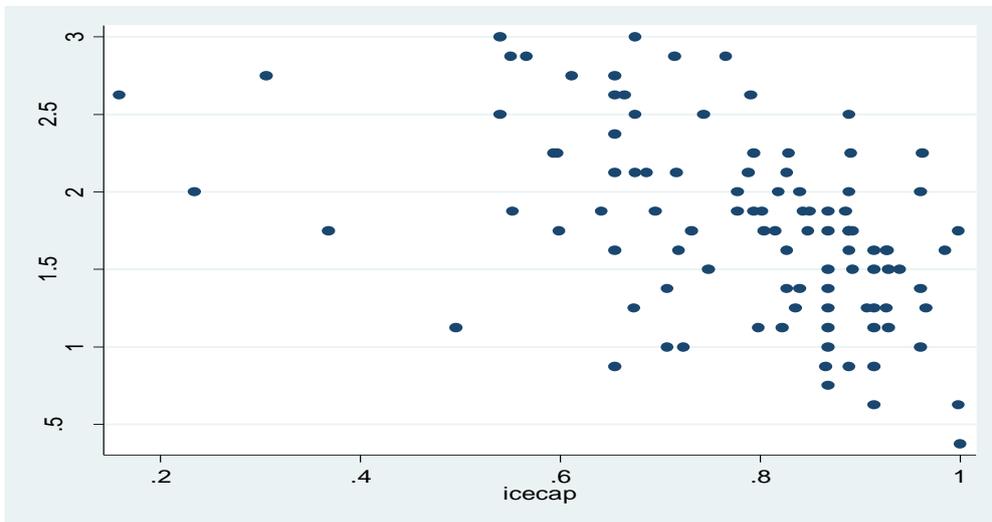
Figure 2. Relationship between ICECAP-O cubic & HAQ-DI: OLS (Model 3 in Table 4&6)
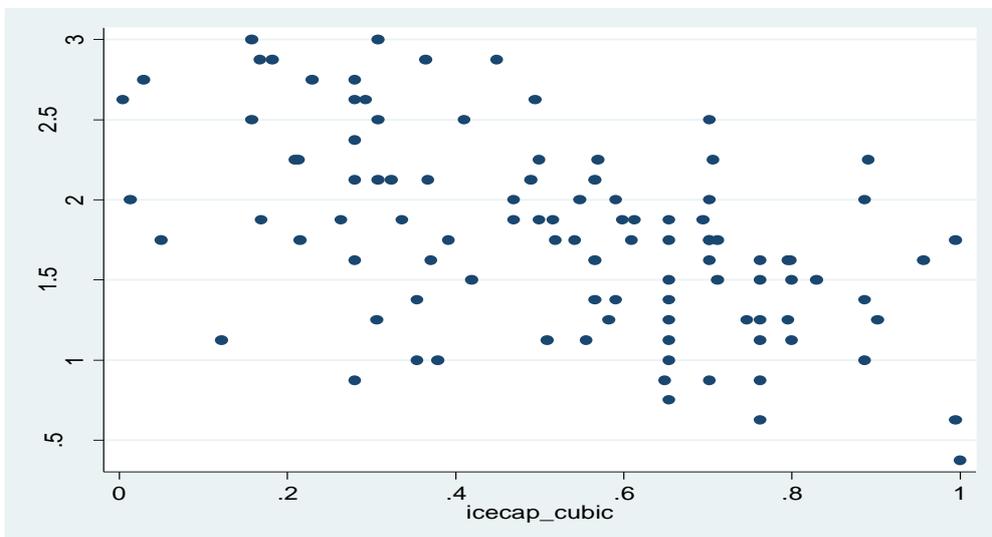
Figure 3 Box-Plots of Attachment Dimension Levels & HAQ-DI Scores
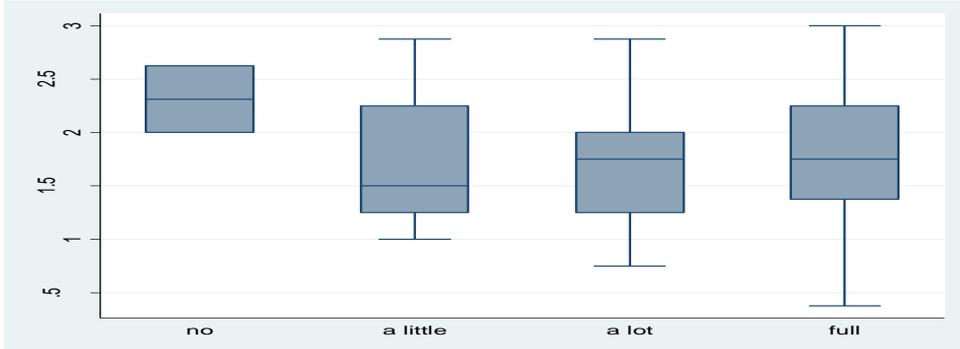
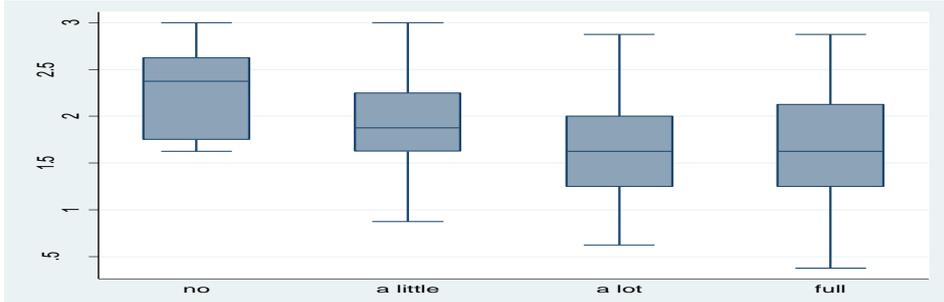Figure 4. Box-Plot of Security Dimension Levels and HAQ-DI Scores

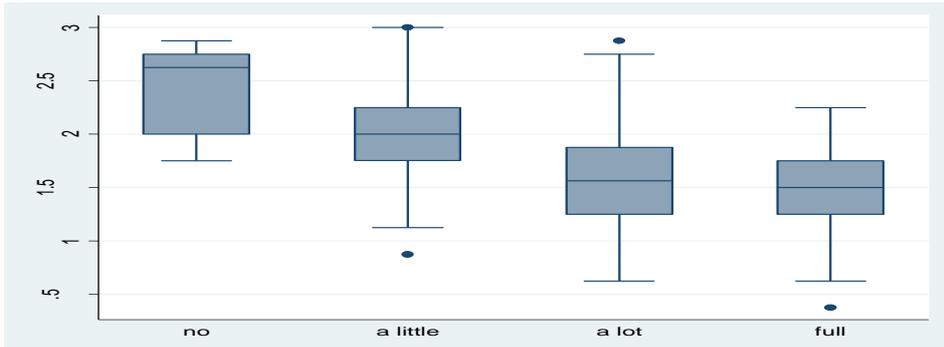Figure 5. Box-Plot of Role Dimension Levels and HAQ-DI Scores

Figure 6. Box-Plot of Enjoyment Dimension Levels and HAQ-DI Scores
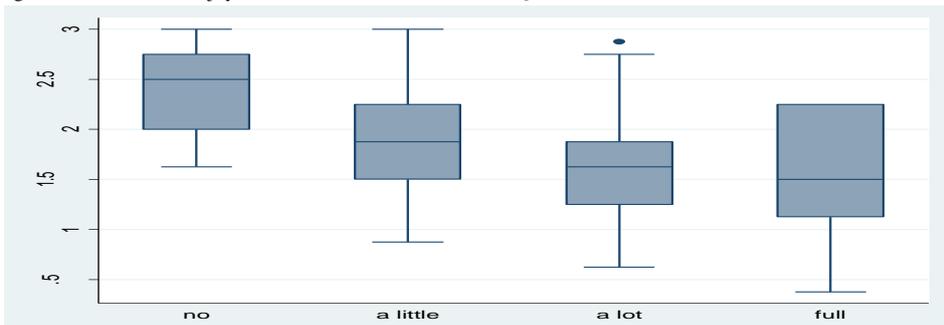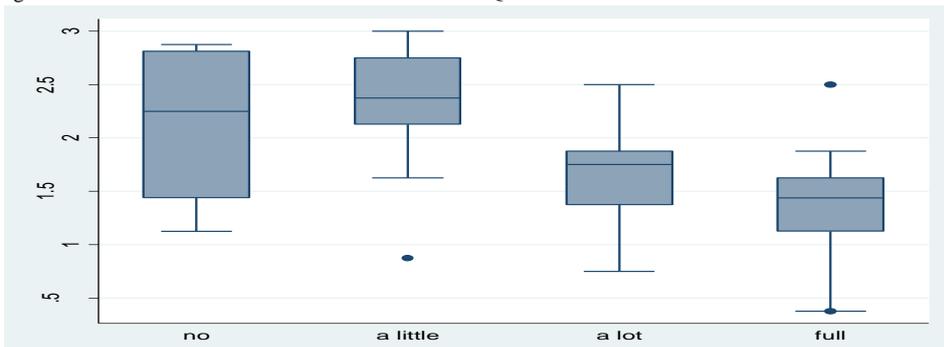
Figure 7 Box-Plots of Control Dimension Levels and HAQ-DI Scores

Appendix I: The ICECAP-O Questionnaire (Coast et al. 2008c): About your quality of life

1. Love and Friendship

I can have all of the love and friendship that I want	4
I can have a lot of the love and friendship that I want	3
I can have a little of the love and friendship that I want	2
I cannot have any of the love and friendship that I want	1

2. Thinking about the future

I can think about the future without any concern	4
I can think about the future with only a little concern	3
I can only think about the future with some concern	2
I can only think about the future with a lot of concern	1

3. Doing things that make you feel valued

I am able to do all of the things that make me feel valued	4
I am able to do many of the things that make me feel valued	3
I am able to do a few of the things that make me feel valued	2
I am unable to do any of the things that make me feel valued	1

4. Enjoyment and pleasure

I can have all of the enjoyment and pleasure that I want	4
I can have a lot of the enjoyment and pleasure that I want	3
I can have a little of the enjoyment and pleasure that I want	2
I cannot have any of the enjoyment and pleasure that I want	1

5. Independence

I am able to be completely independent	4
I am able to be independent in many things	3
I am able to be independent in a few things	2
I am unable to be at all independent	1

Tick one box only in each section

visit www.icecap.bham.ac.uk to access this questionnaire

Appendix II: The HAQ-DI

(Illustration from Barton et al. 2004)

We are interested in learning how your illness affects your ability to function in daily life. Please feel free to add any comments at the end of this form	Without ANY difficulty Score = 0	With SOME difficulty Score = 1	With MUCH difficulty Score = 2	Unable to do Score = 3
PLEASE TICK THE ONE RESPONSE WHICH BEST DESCRIBES YOUR USUAL ABILITIES OVER THE PAST WEEK				
1. DRESSING & GROOMING - Are you able to:				
Dress yourself including tying shoelaces and doing buttons?				
Shampoo your hair?				
2. RISING - Are you able to:				
Stand up from an armless straight chair?				
Get in and out of bed?				
3. EATING - Are you able to:				
Cut your meat?				
Lift a cup or glass to your mouth?				
Open a new carton of milk (or soap powder)?				
4. WALKING - Are you able to:				
Walk outdoors on flat ground?				
Climb up five steps?				
5. HYGIENE - Are you able to:				
Wash and dry your entire body?				
Take a bath?				
Get on and off the toilet?				
6.REACH - Are you able to:				
Reach and get a 5 lb object (e.g. a bag of potatoes) from above your head?				
Bend down to pick up clothing from the floor?				
7. GRIP - Are you able to:				
Open car doors?				
Open jars which have been previously opened?				
Turn taps on and off?				
8. ACTIVITIES - Are you able to:				
Run errands and shop?				
Get in and out of a car?				
Do chores such as vacuuming, housework or light gardening?				

How the HAQ-DI is calculated:

Add the highest score for each of the eight categories together, before dividing by eight to get a score between 0-3.

22 polar questions relating to aids, devices & help required are also asked (not presented here).

For each category where such help is required, a minimum score of 2 is recorded for that category.

Appendix III:

Example of how the HAQ-DI was calculated

(Physical function questions from WOMAC, HAQ-DI & SF-36 – Pollard et al. (2009))

e.g.   Individual A

Degree of Difficulty	*Questionnaire*	None	Mild	Moderate	Severe	Extreme	Score
HAQ Score		0	1	2	3	4	0-4
1. Dressing & Grooming:							
• Dressing yourself	HAQ-DI			X			
• Putting on/off shoes	WOMAC			X			2
• Washing your hair	HAQ-DI			X			
2. Rising:							
• Rising from sitting	WOMAC				X		3
• Rising from bed	WOMAC			X			
• Getting into bed	HAQ-DI			X			
3. Eating:							
N/A							-
4. Walking:							
• Short distances	SF-36				X		3
• Climbing stairs	SF-36			X			
5. Hygiene							
• Washing and drying	HAQ-DI			X			
• Getting in/out of bath	WOMAC				X		3
• Getting on/off toilet	WOMAC			X			
6. Reach							
• Bending to floor	WOMAC				X		3
7. Grip							
N/A							-
8. Activities							
• Going shopping	WOMAC			X			
• Getting in/out of car	WOMAC			X			2
• Light domestic duties	WOMAC			X			
Total Score (0-24)							**16**
Average (Total/6)							**2.67**
HAQ_DI (Average*0.75)							**2**

Once HAQ-DI scores were calculated across the different functioning areas on the 5 point (0-4) WOMAC scale, the total average score was multiplied by 0.75 so that scores correspond with HAQ-DI scores on the 4 point (0-3) HAQ-DI scale used in the BRAM model.