WORK IN PROGRESS – PLEASE DO NOT CIRCULATE OR CITE

RELATIONSHIP BETWEEN THE EUROQOL-5D AND BARTHEL INDEX - EXAMINING THE USE OF PROXY OUTCOME MEASURES FOR OLDER PEOPLE

Billingsley Kaambwa[1]
Stirling Bryan[1]
Pelham Barton[1]
Hilda Parker[2]
Graham Martin[3]

1- Health Economics Facility, Health Services Management Centre, University of Birmingham, Park House, 40 Edgbaston Park Road, Birmingham, B15 2RT

2 - Formerly Nuffield Community Care Studies Unit, University of Leicester now Division of Primary Care, University of Nottingham, Graduate Medical School, Derby City General Hospital, Uttoxeter Road, Derby, DE22

3- Formerly Nuffield Community Care Studies Unit, University of Leicester now IGBiS, University of Nottingham, NG7 2RD

Paper presented at the Health Economists' Study Group Meeting,
University of York, July 2006

Abstract

Purpose

Intermediate care (IC) of older people is a key component of UK government health policy. Evidence on outcomes associated with IC is scarce. In many instances, health related quality of life (HRQoL) outcome measures are not available mainly because older people are physically or mentally not able to (self-) report their HRQoL. The resulting missing outcome values may lead to biased statistical results. Proxy outcome measures may help but their suitability as proxies needs to be tested and verified. The aim of this study was to examine the direct relationship between a conventional clinical scale of functional status that is suited for proxy-assessment (Barthel Index (BI)) and a measure of HRQoL (EuroQoL EQ5D (EQ-5D))

Methods

We considered outcome measures from a total of 964 IC patients participating in the national evaluation of IC for older people in the UK. To test the relationship between the BI and the EQ-5D, we used tests of independence, multiple comparison tests (Dwass-Steel-Critchlow-Flinger tests), Powell's censored least absolute deviations (CLAD) regression, correlation coefficients, and the rank-invariant method of parallel reliability

Results

We established a plausible and significant relationship between the measures. There was evidence of the ability of the BI to predict EQ-5D scores. This relationship was valid for all IC types and settings as well as across all diagnostic groups in the sample.

Conclusions

There is considerable documentation supporting the construct validity of using the EQ-5D in geriatric care and these results add to the evidence that BI scores may be used as proxy for HRQoL in case of missing values on EQ-5D.

Acknowledgements

We would like to acknowledge, with gratitude, the enormous support we have received from colleagues in Leicester and Sheffield. We are also thankful to the intermediate care-coordinators and the staff from the five case-study sites that provided the quantitative data and clarified follow-up questions. This project was sponsored jointly by the UK Medical Research Council and the UK Department of Health Policy Research Programme.

Introduction

Health related quality of life (HRQoL) and functional independence of older people are important outcome measures for evaluating efficacy and effectiveness of therapeutic interventions[1-13]. The former gives an indication of an individual's health status while the latter permits an understanding of how limited individuals are by their functional status. In many well designed evaluation studies on older people, HRQoL data are collected using an appropriate instrument that will allow economic analyses such as cost-effectiveness assessments to be undertaken. One such instrument is the EuroQol EQ-5D (EQ-5D) whose construct validity when used on populations of older people has been well documented.[1,2,6,8] However, it is not always possible to collect EQ-5D data when conducting evaluations of services for older people. Some older people are physically or mentally not able to provide self-report HRQoL data and so the issue is a missing data problem. The important concern then is one of avoiding the bias that might result from simply ignoring the patients with missing data. Sometimes data on HRQoL are simply not available at all because they have not been collected. This tends to be particularly the case when researchers are reliant on routinely collected data for service evaluations. In addition, older people may receive care from different service providers who in turn do not always use the same instruments to document HRQoL. These scenarios therefore highlight the general limitations of relying on just one instrument. An important question to address is whether or not it would still be possible to conduct meaningful economic analyses when information on HRQoL is not available. The use of measures of functional independence or activities of daily living (ADL) as proxy for HRQoL in case of missing values on the EQ-5D has been suggested.[11] But proxy outcome measures may help only if their suitability as proxies is tested and verified. The aim of this study was to examine the direct relationship between a conventional clinical scale of functional independence that is suited for proxy-assessment (Barthel Index (BI)) and the EQ-5D. To do this, the paper examines three main issues:

- The construct validity of each instrument
- Ability of the BI to predict EQ-5D scores
- Parallel reliability between the dimensions of the EQ-5D and the BI

Outcome measures

A number of instruments have been used to measure the HRQoL for older people in care. But prominent among them is the EQ-5D which is a generic quality of life instrument. The EQ-5D is comprised of five dimensions of health: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. There are three levels of impairment in each domain: no, some/moderate, and extreme problems in the relevant dimension of health. Using these responses, the EQ-5D is able to distinguish between 243 states of health.[14,15]. After normalisation, EQ-5D scores have a range of -0.59 to 1: the maximum score of 1 represents perfect health and a score of 0 represents death. Scores less than 1 represent health states that are worse than death[14, 16-19]. Its generic nature makes it comparable across patient populations.

In terms of evaluating functional independence, many measures have been used in geriatric care including the BI. The BI is a functional assessment instrument of activities of daily living (ADL) instrument that has also been widely used to measure health outcomes in populations of older people.[5-8] Further, the BI has been recommended for scientific research purposes and also for use as a proxy for other outcome measures.[20] To measure a person's level of functional independence, the BI uses 10 items, with each item carrying different weights[21]. Two items are rated on a two-point scale of 0 and 5, six on a three-point scale of 0, 5 and 10 and the last two items are rated on a four-point scale of 0, 5, 10 and 15. This resultant scale is from 0 to 100. The scales used in this study have been divided by 5 and therefore range from 0 to 20. The higher the score recorded for an item, the greater the level of independence. The reliability, sensitivity and suitability for proxy-assessment of the BI has been shown elsewhere[21-23]

Research Methods

Data collection

We obtained data from five case study sites that participated in a national evaluation of costs and outcomes of intermediate care (IC) in the UK[24]. IC broadly refers to services that seek to prevent inappropriate hospital admissions, facilitate supported discharges for patients who no longer need to be in hospital, and avert premature admission to long-term care. The 5 sites were chosen according to the following criteria:

- Provided a range of intermediate care services operational for at least 2-3 years;
- Had reasonable throughput into the intermediate care system (at least 1000 cases per annum);
- Were a mix of urban and rural sites;
- Had senior management support for the collection of routine data by services themselves;
- Had clinical and managerial support for participation in the national evaluation.

Staff working for the IC services collected the quantitative data according to protocols set out by the evaluation team. For all IC admissions over a defined period, service staff completed study proformas, with their patients, at the point of entry to the service. They completed discharge questions on the day of discharge, transfer or as soon as possible following end of service provision. Data collection in all five sites continued for approximately seven months. We collected data on patient characteristics, descriptors of the intermediate care services, and intermediate care-related services' descriptors. These data included the two health outcome measures: the EQ-5D and the BI

Statistical Analyses

In order to provide a description of the sample of IC episodes for which data were available, descriptive summary statistics were generated for all the main variables. Tables reporting frequencies and two measures of central tendency (mean & median) were produced. Standard deviations (SD) and the interquartile ranges (IQR) for the mean and median, respectively, were also calculated.

To assess the construct validity of the two instruments, the mean EQ-5D scores were compared against BI categories. As a result of the non-normality of the two measures, the Kruskal-Wallis test[25] was applied to examine the mean difference in the EQ-5D scores across the BI categories. The BI categories were adapted from Wade et al's[26] classification. Five BI categories were considered: independent (BI = 20); mild (BI = 15-19); moderate (BI = 10-14); severe (BI = 5-9); and very severe (BI = 0-4). Differences in the mean of the EQ-5D were also evaluated across three IC functions: acute admission; supported discharge and other IC function. As a follow-up to the Kruskal-Wallis, multiple comparison tests were employed to evaluate which pairs of BI categories and IC functions were significantly different. The Dwass-Steel-Critchlow-Fligner tests[10] were applied to the five categories of the BI and three categories of the IC function

Regression analysis was used to examine the ability of the BI to predict EQ-5D scores. EQ-5D has an upper bound of 1.0 and so regressions using Powell's censored least absolute deviations (CLAD) estimator were used to take the ceiling effect into account.[27,28] The intra-class correlations (ICC) and the Wilcoxon Signed Rank Sum Tests[24] were used to measure the level of correlation and independence, respectively, between the observed and predicted values of the EQ-5D

Examining the level of agreement between individual items, rather than the composite scores, of the two measures gives an indication of the level of interchangeabiltiy between scales. The rank-invariant method of parallel reliability was used to compare all the items in the EQ-5D against each of the dimensions in the BI.[9]The level of agreement in the ordering of paired assessments was measured with the use of contingency tables and a coefficient of monotonic agreement (CMA). A CMA score of 1 indicated perfect agreement while that of 0 depicts complete lack of agreement. Fuller details of the algorithm used to calculate the CMA can be found elsewhere.[9] Figure 1 illustrates this method by looking at the relationship between the self-care and bowels items of the EQ-5D and BI at admission, respectively. The marginal distributions for the BI assessment bowels activity are 17, 44 and 580. The implication for this is that the majority of the 641 individuals (i.e. 580 = 90%) are classified as BI level 2. Of these, 55 represent EQ-5D level 3 implying that these individuals are rated as both EQ-5D 3 and BI 2. The EQ-5D items selected for comparisons against BI items selected according to apriori hypotheses are presented in Table 1. For just this part of the analysis, only 641 such pairs that had complete data at both admission and discharge were used in these comparisons. As for the comparisons that did not have the same number of categories, theoretical cut-off levels as defined by the marginal distributions of the points were determined.[9]

STATA version 8.2[29] was used for all statistical analysis

Results

Demographic characteristics

A sample of 964 IC clients was included in the analysis (Table 2). The mean age was 79.4 and the majority (70%) were female. At the time of inception into an IC service, 54% were living alone. The mean EQ-5D and BI scores at admission were 0.42 and 14.8, respectively. The corresponding mean scores at discharge for these two measures were 0.61 and 16.7. Most IC schemes were performing either an acute admission (55%) or supported discharge (42%) function.

Construct validity of the BI and EQ-5D

Table 3 compares the mean scores of the EQ-5D across the five BI categories. The results show a consistent pattern: very severe functional status (BI = 0-4) is associated with the lowest EQ-5D mean score of -0.20 and very functionally independent status (BI = 20) corresponds to the highest EQ-5D mean score. This relationship holds for data at admission, discharge and for the combined admission and discharge data. Therefore, the EQ-5D differentiates between the five levels of functional status. Table 4 compares the mean scores of both the EQ-5D and BI, across the three IC functions and shows a consistent pattern as well in the mean scores across the IC functions. For instance at admission, patients in IC schemes performing an admission avoidance function had the lowest functional status (mean BI = 13.8) and HRQoL (mean EQ-5D = 0.32) while those in facilities performing a supported discharge function had the highest mean scores (BI = 15.6; EQ-5D = 0.50). The same pattern can be observed when considering data at discharge as well as when using the combined admission and discharge data.

Tables 3 and 4 also show the results of the non-parametric tests of independence. The Kruskal-Wallis test examined whether there was a statistically significant difference in the mean EQ-5D score across the BI categories and also in the mean BI and EQ-5D scores across the IC functions. The results indicate that there were significant differences in the mean EQ-5D across the BI categories at admission (χ^2 = 190.3, *df* = 4, *p*<0.001); at discharge (χ^2 = 144.2, *df* = 4, *p*<0.001) and for the combined admission and discharge data (χ^2 = 393.3, *df* = 4, *p*<0.001). Similar significant results were found across the IC function categories at admission (EQ-5D: χ^2 = 67.1, *df* = 2, *p*<0.001; BI: χ^2 = 46.8, *df* = 2, *p*<0.001); at discharge (EQ-5D: χ^2 = 17.1, *df* = 2, *p*<0.001; BI: χ^2 = 27.0, *df* = 2, *p*<0.001) and for the combined admission and discharge data (EQ-5D: χ^2 = 79.6, *df* = 2, *p*<0.001; BI: χ^2 = 73.2, *df* = 2, *p*<0.001).

As a follow-up to the Kruskal-Wallis test, multiple comparison tests (Dwass-Steel-Critchlow-Fligner tests) were employed. There was statistically significant difference (p<0.001) in the mean EQ-5D across comparisons of all the BI categories (Table 5). Similar results were obtained from the multiple comparison tests assessing whether the mean EQ-5D and BI scores across comparisons of all IC functions were statistically significant (Table 6). The only exception was the pair of 'supported discharge vs Other IC Functions'. This means that both the EQ-5D and BI completely lacked the ability to differentiate between these two IC functions.

Regression results – Predictive ability of the BI

CLAD regression results reported in Table 7 affirm the association of low (high) BI scores with very low (high) EQ-5D which is an indicator of comparable sensitivity of the two measures. At admission, a totally dependent individual (BI=0) has a corresponding EQ-5D score of -0.24, which is worse than death while a BI score of 4.8 is equivalent to death (EQ-5D=0). A totally independent individual (BI = 20) has a corresponding score of 0.76 on the EQ-5D. This is comparable to the reference value of the general population in the UK for ages 65-80.[17] The top EQ-5D score was 0.79 at discharge and also for the combined admission and discharge data. The coefficient of determination were however low for all three equations. At most, the regressions only explained about 15% of the variation in the dependent variable (EQ-5D). Inclusion of other independent variables in the equations did not alter this significantly. Predicted values of the EQ-5D were requested from the CLAD regression and these are shown in Table 8. Tests of independence (Wilcoxon rank sum tests) indicate that the mean values of the observed and predicted EQ-5D scores do not differ significantly from each other across all of the BI categories (p>0.05). The scores differ only for the very severe (BI = 0 -4) group. Using ICCs as measures of association reveals that there is correlation between the observed and predicted values across all BI categories except again for the 'very severe' group. A reason for this may be the very large CI in relation to the other categories. The ICC is however low for all the categories but similar to that of a comparable study[11]

Parallel reliability of the EQ-5D and BI items

Analysing the empirical agreement between items on the EQ-5D and those on the BI sheds light on the level of interchangeabiltiy of scales. A coefficient showing the level of agreement in the ordering of paired assessments (CMA) was calculated from the contingency tables. Movement from the lower left to the upper right of the corner of the tables indicates increased level of HRQoL and functional ability. The category distributions of the EQ-5D and BI appear as marginal frequencies of the contingency tables. Table 9 shows the level of agreement, as depicted by the CMA, for selected pairs of comparison, according to apriori hypotheses. Complete results for all comparisons are shown in Tables 10-12. The results show mild agreement (60-73%) between the EQ-5D and BI items at admission, discharge and also for the combined admissions and discharge data. Only two pairings of items (BI Bowels & EQ-5D self-care/usual activities and BI Transfers & EQ-5D mobility) had high levels of agreement (74-83%).

4. Discussion

Both the EQ-5D and BI have been independently validated as reliable instruments for use with older people.[1, 2,5,12] Some studies have also addressed the use of the BI as a proxy for HRQoL[11,30] Our results have shown that there is a positive relationship between the two measures: lower mean scores on the BI correspond to lower mean scores on the EQ-5D. In addition, the regression results have revealed that the BI has the ability to correctly predict EQ-5D scores This finding is useful for circumstances where data on the HRQoL is missing. A limitation here though is that the coefficient of determination in the regression model was much lower than that in a comparable study.[11] What needs to be borne in mind however is that the latter study focussed on stroke patients for whom the use of the BI was more appropriate. Analysis of subsection of stroke patients in this study (results not reported) yielded identical results to this earlier piece of research. A further limitation may have been the adoption of the method used to condense the BI score into broad categories. It remains to be seen if different classifications may yield different results than those we found. The effect of the different ways of arriving at a composite score for the two measures has not been investigated and may be important for future study: the EQ-5D uses a tariff while the BI just adds up the scores for the items. The results from the rank-invariant method of parallel reliability explain why we do not have extremely high concordance between the two measures. The results reflect the difficulty that exits with direct and perfect mapping of one instrument to another. There was mild to strong agreement for most of the pairs and this may reflect the fact that these measures were designed to place emphasis on different aspects of outcomes in the first place.

To conclude, this study has shown that a mild to strong concordance exists between the two measures used for measuring outcomes among older people adding to the evidence of the suitability of using the BI as a proxy measure for the EQ-5D. Though not a perfect relationship, the relationship depicted is still useful in understanding the correlation between HRQoL and functional status or independence. Despite the shortcomings associated with the use of measures of ADL such as the BI as proxies for HRQoL, one should not abandon them but rather evaluate them using appropriate criteria that will produce better fits. Ultimately, the choice of an outcome measure largely depends on who wants the data and for what purpose it is needed

References

1. Coast J, Peters TJ, Richards SH, Gunnell DJ. Use of the EuroQoL among elderly acute care patients. *Qual Life Res* 1998; **7**(1):1-10.

2. Brazier JE, Walters SJ, Nicholl JP, Kohler B. Using the SF-36 and EuroQoL on an elderly population. *Qual Life Res* 1996; **5**: 195–204.

3. King RB. Quality of Life after Stroke. *Stroke* 1996; **27**:1467-1472.

4. Griffiths P, Harris R, Richardson G, Hallett N, Heard S, Wilson-Barnett J. Substitution of a nursing-led inpatient unit for acute services: randomized controlled trial of outcomes and cost of nursing-led intermediate care. *Age Ageing* 2001; **30**: 483-488.

5. Yohannes A, Roomi J, Waters K and M Connolly. A comparison of the Barthel index and Nottingham extended activities of daily living scale in the assessment of disability in chronic airflow limitation in old age. *Age Ageing* 1998; **27**: 369-374.

6. Van den Bos G, Triemstra AHM. Quality of life as an instrument for need assessment and outcome assessment of health care in chronic patients. *Qual Health Care* 1999; **8**:247–252.

7. Davies P. Social approaches to health outcomes. In Macbeth HM (ed). *Health Outcomes: Biological, Social, and Economic Perspectives.* New York: Oxford University press, 1996; 100-105.

8. Lyons RA, Crome P, Monaghan S, Killalea D and Daley JA. Health status and disability among elderly people in three UK districts. *Age and Ageing* 1997; **26**: 203-209.

9. Gosman-Hedstrom G, Svensson E. Parallel reliability of the Functional Independence Measure and the Barthel ADL index. *Disability and Rehabilitation* 2000; **22**(16): 702-715

10. Kwon S, Hartzema AG, Duncan PW, Lai S. Relationship Among the Barthel Index, the Functional Independence Measure and the Modified Rankin Scale. *Stroke* 2004; **35**: 918-923

11. van Exel NJA, Scholte op Reimer WJM, Koopmanschap MA. Assessment of post-stroke quality of life in cost-effectiveness studies: The usefulness of the Barthel Index and the EuroQoL-5D. *Quality of Life Research* 2004; **13**; 427-433

12. Sainsbury A, Seebass G, Bansal A, Young JB. Reliability of the Barthel Index when used with older people. *Age and Ageing* 2005; **34**: 228-232

13. Gauggel S, Bocker M, Heinaemann AW, Lammler G, Borchelt M, Steinhagen-Thiessen E. Patient-Staff Agreement on Barthel Index Scores at Admission and Discharge in a Sample of Elderly Stroke Patients. *Rehabilitation Psychology*; **49** (1): 21-27

14. Dolan P. Modeling Valuations for EuroQol Health States. *Med Care* 1997; **35**(11):1095-1108.

15. Bowling A. *Measuring health: a review of quality of life measurement scales*. Maidenhead: Open University Press, 2005

16. Royal College of Physicians of London. *Standardised assessment scales for elderly people*. London: Royal College of Physicians of London, 1992

17. Kind P, Hardman G, Macran S. UK population norms for EQ-5D: Discussion paper 172. New York, UK: Centre for Health Economics, The University of York, 1999.

18. Murphy R, Sackley CM, Miller P, Harwood RH. Effect of experience of severe stroke on subjective valuations of quality of life after stroke. *J Neurol eurosurg Psychiatry* 2001; **70**: 679–681.

19. Post PN, Stiggelbout AM, Wakker PP. The utility of health states after stroke. *Stroke* 2001; **32**: 1425–1429.

20. Wilkinson PR, Wolfe CD, Warburton FG, et al. Longer term quality of life and outcome in stroke patients: Is the Barthel Index alone an adequate measure of outcome? *Qual Health Care* 1997; **6**(3): 125–130.

21. Mahoney FI, Barthel DW. Functional evaluation: The Barthel Index. *Maryland State Med J* 1965; **14**: 61–65.

22. Shah S, Vanclay F, Cooper B. Improving the sensitivity of the Barthel Index for stroke rehabilitation. *J Clin Epidemiol* 1989; **42**: 703–709.

23. Wolfe CDA, Taub NA, Woodrow EJ, Burney PGJ. Assessment of scales of disability and handicaps for stroke patients. *Stroke* 1991; **22**: 1242–1244.

24. ICNET Team. A National Evaluation of the Costs and Outcomes of Intermediate Care for Older People: Final Report. Leicester: The University of Leicester, 2006.

25. Ott RL. *An Introduction to Statistical Methods and Data Analysis 4th edition*. Belmont CA: Duxbury Press, 1993

26. Wade DT, Collin C. The Barthel ADL Index: A standard measure of physical disability? *Int Disabil Studies* 1988; **10**: 64–67

27. Chay KY, Powell JL. Semiparametric Censored Regression Models. *Journal of Economic Perspectives*; **15**(4): 29-42.

28. Clarke P, Gray A, Holman R. Estimating Utility Values for Health States of Type 2 Diabetic Patients Using the EQ-5D (UKPDS 62). *Med Decis Making* 2002; **22**: 340-349

29. StataCorp LP. Intercooled Stata 8.2 for Windows. Texas: College Station, 2004.

30. Coons SJ, Rao S, Keininger DL, Hays RD. A Comparative Review of Generic Quality-of-Life Instruments. *Pharmacoeconomics* 2000; **17**(1):13-35

Figure 1: Barthel Bowels & EQ Self-care

		BI			
		0	1	2	*f*
EQ-5D	3	6	10	55	71
	2	6	28	334	368
	1	5	6	191	202
	f	17	44	580	641

Figure 2: BI & EQ-5D at admission

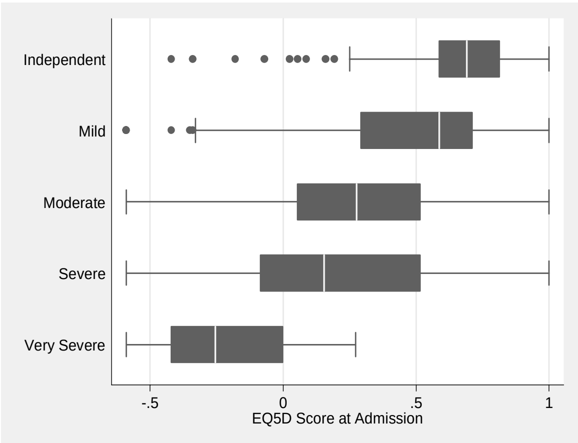

Figure 3: BI & EQ-5D at discharge

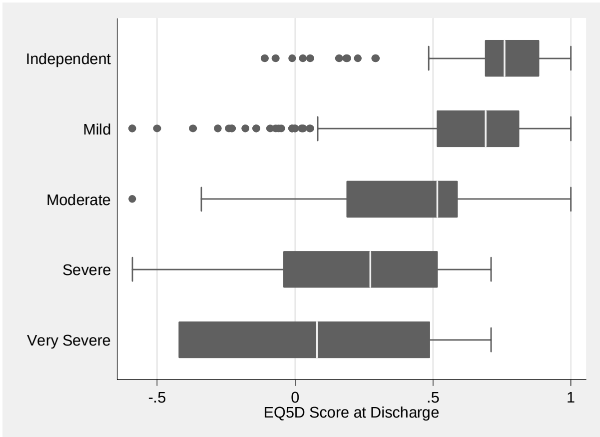

Table 1: Items compared in the Barthel ADL Index and the EuroQol EQ-5D (n=641)

Barthel	**EQ-5D**
Feeding	Self-care/Usual activities
Bathing	Self-care
Grooming	Self-care
Dressing	Self-care
Bowels	Self-care/Usual activities
Bladder	Self-care/Usual activities
Toilet Use	Self-care
Transfers	Mobility
Mobility	Mobility
Stairs	Mobility
	EQ-5D items not included
	Pain/discomfort
	Anxiety/depression

Table 2: Sample demographic and baseline characteristics

Age - Mean (SD)	79.4 (10.34)
% Male	30
% Living Alone	54
EQ-5D at Admission - Mean (SD)	0.42 (0.36)
EQ-5D at Discharge - Mean (SD)	0.61 (0.31)
Barthel at Admission - Mean (SD)	14.8 (4.22)
Barthel at Discharge - Mean (SD)	16.7 (3.85)
% Admission Avoidance	55
% Supported Discharge	42
% Residential Intermediate Care	18

Table 3: Mean EQ-5D across different BI Categories

	At Admission N=964		At Discharge N=728		Combined N=1692	
BI Category*	Mean EQ-5D	95% CI	Mean EQ-5D	95% CI	Mean EQ-5D	95% CI
Independent (BI=20)	0.65	0.59; 0.70	0.76	0.72; 0.79	0.72	0.69; 0.74
Mild (B1= 15-19)	0.51	0.48; 0.54	0.61	0.59; 0.64	0.56	0.54; 0.58
Moderate (B1= 10-14)	0.28	0.24; 0.32	0.39	0.33; 0.46	0.31	0.27; 0.34
Severe (B1= 5-9)	0.18	0.10; 0.26	0.18	0.02; 0.34	0.18	0.11; 0.25
Very Severe (B1= 0-4)	-0.20	-0.38;-0.02	0.09	-0.32; 0.51	-0.09	-0.27; 0.08

Means per BI category differ significantly from each other (Kruskall Wallis; p<0.001)
** BI categorisation adopted from Wade et al (1988)*

Table 4: Mean Barthel & EQ-5D across Service Functions

	Service Function					
	Admission Avoidance N=930		Supported Discharge N=711		Other IC Function N=21	
	Mean	95% CI	Mean	95% CI	Mean	95% CI
BI – Admission*	13.8	13.5; 14.2	15.6	15.3; 15.9	15.2	14.1; 16.4
BI – Discharge*	16.0	15.6; 16.4	17.2	16.9; 17.5	16.3	15.1; 17.5
BI – Combined*	14.7	14.4; 15.0	16.4	16.1; 16.6	15.7	14.9; 16.6
EQ-5D-Admission*	0.32	0.28; 0.35	0.50	0.47; 0.53	0.49	0.39; 0.58
EQ-5D-Disharge*	0.56	0.52; 0.59	0.65	0.62; 0.67	0.60	0.51; 0.69
EQ-5D-Combined*	0.41	0.39; 0.44	0.57	0.55; 0.58	0.54	0.47; 0.60

**Means per Service Function differ significantly from each other (Kruskall Wallis; p<0.001)*

Table 5: Multiple Comparison Tests – Dwass-Steel-Critchlow-Flinger

	EQ-5D Mean Difference
Independent vs Mild	-0.16*
Independent vs Moderate	-0.41*
Independent vs Severe	-0.54*
Independent vs Very Severe	-0.81*
Mild vs Moderate	-0.25*
Mild vs Severe	-0.38*
Mild vs Very Severe	-0.65*
Moderate vs Severe	-0.13*
Moderate vs Very Severe	-0.40*
Severe vs Very Severe	-0.27*
All levels	KW =394.26*

** p<0.001; KW shows Kruskal Wallis test statistic ;*

Table 6: Multiple Comparison Tests – Dwass-Steel-Critchlow-Flinger

	EQ-5D Mean Difference	Barthel Mean Difference
Acute Admission vs Supported Discharge	0.15*	1.63*
Acute Admission vs Other IC Functions	0.12*	1.00*
Supported Discharge vs Other IC Functions	-0.03 (0.73)	-0.63 (0.36)
All Levels	KW=79.78*	KW=74.07*

** p<0.001; KW shows Kruskal Wallis test statistic*

Table 7: CLAD Regression *Results*

	At Admission N=964		At Discharge N=728		Combined N=1692	
Variable	Coeff.[a]	95% CI	Coeff.[b]	95% CI	Coeff.[c]	95% CI
Constant	-0.24	-0.36;-0.12	-0.008	-0.20;0.19	-0.21	-0.29;-0.13
BI	0.05	0.04;0.05	0.04	0.03; 0.05	0.05	0.04;0.05

Dependent Variable; EQ-5D [a] *R^2 = 0.14;* [b] *R^2 = 0.12* [c] *R^2 = 0.15* *$p<0.001$*

Table 8: Observed and predicted EQ-5D values

	Observed EQ-5D		Predicted EQ-5D		ICC
BI Category	Mean EQ-5D (SD)	95% CI	Mean EQ-5D(SD)	95%CI	
Independent (BI=20)	0.72 (0.01)	0.69;0.74	0.76 (0)	NA	NA
Mild (B1= 15-19)	0.56 (0.01)	0.54;0.58	0.62 (0.002)	0.62;0.63	0.28*
Moderate (B1= 10-14)	0.31 (0.02)	0.27;0.34	0.38 (0.004)	0.37;0.39	0.15*
Severe (B1= 5-9)	0.18 (0.03)	0.11;0.25	0.15 (0.006)	0.14;0.16	0.25*
Very Severe (B1= 0-4)	-0.09 (0.08)	-0.27;0.08	-0.1 (0.016)	-0.13;-0.06	0.28
Total	0.50 (0.01)	0.49;0.52	0.56 (0.004)	0.55;0.57	0.52*

Means per BI category do not differ significantly from each other (Wilcoxon Test; $p>0.05$)
*ICC = intra-class correlation (Spearman's p, two-tailed). *$p < 0.001$*

Table 9: Parallel Reliability between BI & EQ-5D – Coefficient of Monotonic Agreement (CMA)

BI vs EQ-5D Items	At Admission N=641	At Discharge N=641	Combined N=1282
BI Feeding vs EQ-5D Self-care	0.56	0.65	0.59
BI Feeding vs EQ-5D Usual activities	0.63	0.72	0.65
BI Bathing vs EQ-5D Self-care	0.59	0.43	0.45
BI Grooming vs EQ-5D Self-care	0.52	0.41	0.42
BI Dressing vs EQ-5D Self-care	0.23	0.08	0.12
BI Bowels vs EQ-5D Self-care	0.77	0.75	0.76
BI Bowels vs EQ-5D Usual activities	0.83	0.80	0.82
BI Bladder vs EQ-5D Self-care	0.62	0.56	0.60
BI Bladder vs EQ-5D Usual activities	0.67	0.67	0.66
BI Toilet vs EQ-5D Self-care	0.63	0.53	0.63
BI Transfer vs EQ-5D Mobility	0.74	0.66	0.73
BI Mobility vs EQ-5D Mobility	0.57	0.64	0.57
BI Stairs vs EQ-5D Mobility	0.50	0.46	0.44

Table 10: Monotonic Agreement between all EQ5D and Barthel dimensions at admission

		EQ5D Dimensions				
		Mobility	Self-care	Usual activities	Pain	Anxiety
Barthel Dimensions	Grooming	0.75	0.52	0.76	0.66	0.52
	Toilet	0.62	0.44	0.48	0.55	0.49
	Feeding	0.71	0.56	0.63	0.68	0.71
	Transfer	0.74	0.59	0.64	0.67	0.60
	Mobility	0.75	0.62	0.65	0.67	0.61
	Dressing	0.64	0.43	0.34	0.47	0.43
	Stairs	0.50	0.39	0.40	0.47	0.47
	Bathing	0.72	0.59	0.70	0.76	0.69
	Bladder	0.73	0.62	0.66	0.69	0.64
	Bowels	0.77	0.77	0.83	0.79	0.77

Table 11: Monotonic Agreement between all EQ5D and Barthel dimensions at discharge

		EQ5D Dimensions				
		Mobility	Self-care	Usual activities	Pain	Anxiety
Barthel Dimensions	Grooming	0.64	0.41	0.72	0.70	0.60
	Toilet	0.67	0.53	0.64	0.78	0.66
	Feeding	0.76	0.65	0.72	0.78	0.76
	Transfer	0.79	0.68	0.80	0.82	0.76
	Mobility	0.78	0.68	0.77	0.82	0.73
	Dressing	0.53	0.08	0.43	0.59	0.48
	Stairs	0.46	0.38	0.42	0.47	0.52
	Bathing	0.58	0.43	0.60	0.64	0.61
	Bladder	0.65	0.56	0.67	0.70	0.65
	Bowels	0.84	0.75	0.08	0.85	0.84

Table 12: Monotonic Agreement between EQ5D and Barthel dimensions for combined admission and discharge data

		EQ5D Dimensions				
		Mobility	Self-care	Usual activities	Pain	Anxiety
Barthel Dimensions	Grooming	0.64	0.51	0.76	0.66	0.52
	Toilet	0.61	0.44	0.48	0.55	0.48
	Feeding	0.72	0.56	0.63	0.68	0.71
	Transfer	0.73	0.59	0.64	0.67	0.60
	Mobility	0.73	0.61	0.65	0.68	0.60
	Dressing	0.51	0.43	0.34	0.47	0.43
	Stairs	0.44	0.39	0.41	0.47	0.47
	Bathing	0.59	0.59	0.70	0.76	0.69
	Bladder	0.68	0.62	0.66	0.69	0.64
	Bowels	0.81	0.77	0.83	0.79	0.77