Full Title: Utility of a three item short form version of the Barthel Index for use in Stroke: systematic review and external validation

Cover Title: Short form Barthel Index

Author List: Rachael L MacIsaac PhD[1]; Myzoon Ali PhD[1]; Martin Taylor-Rowan MSc[1]; Helen Rodgers FRCP[2]; Kennedy R Lees FRCP[1]; Terence J Quinn MD[1]

on behalf of the VISTA Collaborators*

1. Institute of Cardiovascular and Medical Sciences, University of Glasgow

2. Stroke Research Group Institute for Neuroscience and Institute for Ageing, Newcastle University

*VISTA Collaborators listed in Supplementary materials

Author for correspondence: Dr Terry Quinn
Institute of Cardiovascular and Medical Sciences, University of Glasgow
New Lister Building, Glasgow Royal Infirmary
Glasgow, UK G4 0SF
Email:terry.quinn@glasgow.ac.uk
Tel:+44 141 211 8510

Word count:4,998 inclusive (abstract:250)

Tables:

Table 1: Papers describing short forms of the Barthel Index

Table 2: Correlation (Spearman's rho) for each Barthel Index item with total Barthel Index score

Table 3: Correlation of various forms of Barthel Index with other outcomes

Table 4:Association of various forms of Barthel Index at baseline with 90 day outcomes

Figures:

Figure: Items included in short forms of Barthel Index

Journal Subject Codes: Stroke [45] Acute Stroke Syndromes

Keywords:activities of daily living, Barthel Index, item-reduction, stroke

ABSTRACT

Background and purpose:There may be potential to reduce the number of items assessed in the Barthel Index (BI) and shortened versions of the BI have been described. We sought to collate all existing short form BI (SF-BI) and perform a comparative validation using clinical trial data.

Methods:We performed a systematic review across multidisciplinary electronic databases to find all published SF-BI. Our validation used the Virtual International Stroke Trials Archive (VISTA) resource. We describe concurrent validity (agreement of each SF-BI with BI); convergent and divergent validity (agreement of each SF-BI with other outcome measures available in the dataset); predictive validity (association of prognostic factors with SF-BI outcomes) and content validity (item correlation and exploratory factor analyses).

Results:From 3546 titles, we found eight papers describing six differing SF-BI. Using acute trial data (n=8852), internal reliability suggested redundancy in BI (Cronbach α:0.96). Each SF-BI demonstrated a strong correlation with BI, modified Rankin Scale (mRS), National Institutes of Health Stroke Scale (NIHSS) (all $\rho \geq 0.83$; $p < 0.001$). Using rehabilitation trial data (n=332) SF-BI demonstrated modest correlation with quality of life measures SIS and EQ-5D ($\rho \geq 0.50$, $p < 0.001$). Pre-specified prognostic factors were associated with SF-BI outcomes (all $p < 0.001$). Our factor analysis described a three factor structure and item reduction suggested an optimal 3-item SF-BI comprising bladder control, transfer and mobility items in keeping with one of the 3-item SF-BI previously described in the literature.

Conclusion:There is redundancy in the original BI; we have demonstrated internal and external validity of a 3-item SF-BI that should be simple to use.

INTRODUCTION

The Barthel Index (BI) is a ten-item measure of basic activities of daily living (ADL).[1] The BI is the second most commonly used functional assessment scale in stroke trials and the most commonly used ADL assessment in adult rehabilitation.[2,3] BI quantifies ADL in an ordinal, hierarchical scale that ranges from 0-20 or 0-100 depending on the scoring used.[4] BI is recommended as an outcome measure by various professional societies and guidelines.[3] BI has proven prognostic utility[5], it is used in clinical practice to inform rehabilitation and care planning and it is used in research both to describe outcomes and as case-mix adjuster. The BI has proven a useful scale but there is scope for improvement, for example "floor and ceiling" effects of BI scoring are well described.[6] For any assessment there is a trade-off between the time and effort required for testing and the validity of the data acquired[7]. Although administration time for BI assessment is modest, there is still opportunity cost, particularly in busy clinical settings.[8]

Issues with time taken to complete a scale are important to the assessor (longer time spent in assessment gives less time for other clinical activity) and is important to the patient (test burden is a particular issue in the context of acute stroke). These issues will be more apparent in patients with physical, cognitive or communication difficulties, yet this is exactly the population that require robust assessment of function. In the NHS England and Wales National Stroke Audit, completion rate of BI measures was approximately 60%, with lack of time cited as the reason for poor completion.[9] The problem is not unique to assessment of BI, in large registries completion of the modified Rankin Scale (mRS) was around 75% with lesser completion in those with more severe impairments[10]. In a rehabilitation study, completion of the Stroke Impact Scale was limited with potential to bias results[11].

In this situation, the ideal would be a shortened form of the BI that offered prompt assessment without sacrificing clinical properties. The high internal reliability of the BI suggests that certain component items of BI are redundant and there is potential to condense the scale.[6]

We sought to describe and compare properties of published short form versions of BI (SF-BI), using a two stage approach; first systematically searching the literature for SF-BI and then validating and comparing the various forms using an independent dataset.

Methods

Systematic review

Our primary question for the systematic review was:which items are included in short form versions of BI for use in patients in stroke? As the purpose of the search was to find SF-BI we did not perform quantitative summary analyses or quality assessment of primary papers.

We devised a focussed search strategy using validated search terms across multidisciplinary electronic databases. After initial scoping searches, we opted to use a concepts-based approach with search strings based on concepts of "Barthel Index/ADL-assessment" and "short forms/psychometric properties of scales". Search strings were based on MeSH and other controlled vocabulary.(Supplementary materials I)

We searched across three electronic databases (Medline (OVID), Embase (OVID) and Health and Psychosocial Instruments (OVID)) all from inception to December 2015. We used citation searching (backwards searching) and assessed all papers that had cited the index paper (forwards searching).

We included any paper that described a shortened (less than ten items) version of the BI. We limited to studies of patients with stroke or brain injury but operated no restrictions with respect to language, date or study design.

Titles and abstracts generated from the electronic database searches were screened for relevance. Irrelevant titles and abstracts were excluded and full-text articles inspected to determine eligibility. As a test of external validity, we pre-selected two studies[12,13] relevant to the study question from a previous review of BI properties[6] and we assessed whether the search included these studies.

We extracted details of studies meeting inclusion criteria to a pre-specified proforma. We described the items included in the short form, the derivation sample, the method used for item reduction and any validation. We included data in the primary publications and supplementary materials but did not contact study authors for additional detail.

We followed, where appropriate, PRISMA best practice guidance for design, conduct and reporting of this systematic review.[14] All aspects of title selection, assessment and data extraction were performed by two independent researchers trained in systematic review (TQ, MTR). The review protocol was registered with Research Registry (www.researchregistry.com, researchregistry1213).

Validation of Short Form Barthel Index

Validity is the extent to which a rating scale measures what it purports to measure.[15] We used multiple, complementary, approaches to validation of each of the SF-BI identified by literature searching, all pre-specified and described in our protocol. Based on peer review advice, we added a further assessment of divergent validity. For these analyses we utilised the Virtual International Stroke Trial Archive (VISTA) resource. All analyses used SAS version 9.4 (SAS Institute, Cary USA) software.

VISTA is a not-for-profit repository for stroke trial data, containing study quality, anonymized individual patient level data on thousands of participants. These data have been used to investigate novel hypotheses including analyses of stroke assessment scale properties.[16,17] We selected all patient-level data that contained BI along with any other functional outcome measure. Within VISTA we had access to datasets from acute stroke settings and rehabilitation studies. We ensured that studies included in our VISTA datasets had not been used to develop any of the published SF-BI found on literature searching. A-priori we decided to treat data from acute stroke trials and from rehabilitation studies separately, as we believed they may have differing outcome measures and differing case-mix of participants.

We described clinical and demographic features of the acute and rehabilitation datasets. Where data were collected at more than one time-point, we used the time-point that gave largest dataset. We assessed internal consistency of standard (ten item) BI using Cronbach's alpha.

Concurrent, convergent and divergent validity: We described concurrent validity by assessing the agreement (Spearman's rank correlation) of each SF-BI with the standard BI and with the various other short forms. We assessed convergent validity by describing agreement with other functional outcome assessments. After initial scoping of available data, our chosen comparator outcomes for acute data were modified Rankin Scale (mRS) and National Institutes of Health Stroke Scale (NIHSS). For rehabilitation data, we used the health related quality of life tools five domain EuroQOL (EQ-5D and visual analogue scale [VAS]) and the Stroke Impact Scale (SIS). EQ-5D data were transformed into a single index using the Europe-VAS dataset. For divergent validity we described association with an aphasia scale, the Sheffield Screening Test for Acquired Language Disorders. We hypothesised that agreement with SF-BI would be greater for the other activity/impairment

level scales (mRS, NIHSS), less for the quality of life scales (EQ-5D, VAS, SIS) and lowest for the aphasia scale.

Predictive validity: We used ordinal univariate regressions to assess association of each SF-BI with, where data were available, clinical and demographic features known to influence stroke outcome (age, baseline stroke severity, physiological variables, comorbidity, prior stroke and use of thrombolytic therapy). In the first analysis we described cross-sectional association of point change in various SF-BI with clinical and demographic factors known to be associated with outcome. In the second analysis we described odds of a point change in SF-BI associated with unit change in NIHSs or mRS at 90-day follow-up.

Content validity: As a final test we explored the most discriminating BI items in the acute VISTA dataset. We performed exploratory factor analysis to suggest a minimum number of items for a short form and further analyses to determine the optimal items for this short form. Due to the larger sample size, factor analysis was restricted to the acute dataset. We first described correlation, using Spearman's rho, for each BI item relative to the total score. Correlations between individual items were explored to investigate item redundancy and exploratory factor analysis was performed to investigate the underlying structure of the BI. As a final test of content validity we derived a short form from the VISTA data. We used exploratory factor analysis to outline the minimum number of factors needed and then used a stepwise selection process sequential removing the poorest performing individual item (based on correlation with total BI and Cronbach's alpha) and comparing properties to find the three items within VISTA that had optimal properties. We compared the resulting VISTA derived short form (herein referred to as SF-BI VISTA) with the short forms identified from the systematic review.

Results

Systematic review: From 3546 titles, we found eight[12,13,18-23] titles describing six differing short forms of the BI.(PRISMA diagram, Supplementary materials II) Some of the papers were validations of previously described scores[20,19], although it was not always clear in the

text if the SF-BI presented was derivation or validation. The validity of our search was proven as our two pre-specified papers[12,13] were included in the original search results.

The short forms differed in the number of included items, the nature of the items included and in the methodology used for item reduction. The short forms included a variety of BI items and all BI items were included in at least one of the short forms. Ability to perform transfers was a feature in most of the SF-BI, while dressing was included in only one SF-BI.(Figure 1) The short forms described by Bohannon and Ellul required additional computation to assign a total score[13,19] and we added these formulae to correct the score before any of the validation analyses.

The authors used various approaches to derivation of the SF-BI and methods of derivation and validation were not consistently described. One of the papers described a short form with no reference to derivation or validation.[23] Only the 5-item SF-BI described by Hobart[12] and 3-item SF-BI described by Ellul had robust derivation, multi-modal validation and further validation of the unmodified scale in an external dataset.(Table 1)

Validation analyses: The VISTA database had 8852 acute strokes with a recorded measurement of BI at 90 days (919 ICH, 7933 ischaemic), 8493 of whom had a complete BI measurement for day 30. The rehabilitation dataset had 332 participants with a recording of BI at baseline. For these rehabilitation studies, "baseline" assessments were predominantly at four weeks post ictus. The included patients were broadly representative of trial populations, mean age 68.1 years (SD:12.4), n=3943 (44.5%) female for the acute dataset and 65.7 years (SD:11.0), n=107 (32%) female for the rehabilitation dataset. Both populations had prevalent comorbidity for example, ischaemic heart disease and diabetes mellitus.(Supplementary materials III,IV)

There was a spread of BI scores across both datasets, for acute data median BI day 90 was 80 (IQR:60), for rehabilitation data, median BI baseline was 75 (IQR:35). Internal consistency for complete BI was high in the acute dataset, with α:0.95 (BI day 30 and 90). For the rehabilitation dataset α:0.85 (BI baseline).

In both datasets we described concurrent validity as the correlation of each individual BI item with the full scale.(Table 2, Supplementary materials V)

We assessed convergent validity of each SF-BI in our datasets. Agreement of SF-BI with full BI was excellent in both datasets.(Table 3) Each SF-BI showed significant ($p<0.0001$) correlations with all our chosen outcome measures. For acute data correlations with mRS and NIHSS were strong. SF-BI at baseline showed weaker correlation with quality of life measures in the rehabilitation dataset, albeit correlations were roughly equivalent to those seen for full BI. Correlations were strongest for SIS, a measure that includes assessment of ADL, and weakest for the VaS. Correlations with the aphasia measure (divergent validity) were weak.(Table 3)

Our assessment of predictive validity was limited to the acute dataset, due to small numbers of common follow-up assessments in the rehabilitation dataset. On ordinal univariate analyses, several factors known to predict outcome were independently associated ($p<0.0001$) with SF-BI.(Supplementary materials VI) Each SF-BI was independently predictive of mRS at day 90 and NIHSS at day 90 in univariate ordinal regressions.(Supplementary material VII) This association persisted when adjusting for the relevant clinical attributes suggested in univariate analysis (age,sex,stroke type).(Table 4)

As a test of content validity we derived a correlation matrix, between item analyses suggested redundancy with correlations of >0.7 for most individual items.(Supplementary materials V). Exploratory factor analysis identified two independent factors within BI, with a potential third cross-loading factor.(Supplementary materials VIII) Based on this, we derived a 3-item SF-BI and for comparison a 5-item SF-BI. The optimal 3-item scale comprised bladder control, transfer and mobility i.e. the items used in Ellul.[13] The optimal 5-item scale compromised dressing, toileting, transfers, mobility, stairs.

As a post-hoc exploratory analysis we compared the utility of a simple sum of the three items, as used in SF-VISTA, and compared to the scores generated using the formulae suggested by Ellul.[13] In the context of our validation analyses we found that, compared to the Ellul formula, there was no evidence that the simple sum score used in SF-VISTA correlated less well with other BI-derived scales (Table 3) or was less strongly associated with mRS or NIHSS (Table 4).

Discussion

Using systematic review and secondary analyses of existing data we have described the validity of various published short form versions of the BI. Our review of the literature found various SF-BI with differing number of items, differing components included and differing scoring. The derivation and validation of these scales was inconsistently described. However, our independent validation using a large dataset confirmed the potential item redundancy within BI (high internal reliability) and suggested utility of a short form (the form originally described by Bohannon[18,19] and Ellul[13]) comprising three easily assessed variables.

Based on our analyses we would recommend a 3-item SF-BI that assesses bladder control, mobility and transfers. We feel this offers parsimony while still capturing key aspects of ADL. Comparing the existent three, four and five item SF-BI, there was no obvious increase in our measures of validity with increasing number of items. We note that ability to perform transfers appeared in almost all the short forms and suspect that any short form should include this item. The component items of the 3-item SF-BI should be relatively simple to score and our post-hoc analysis suggests that item scored can be added to give a total score without the need for additional calculations.

We chose to validate existing SF-BI rather than focus on creating our own de-novo SF-BI. A-priori we suspected that various SF-BI would be available and we recognise the difficulty of establishing a novel assessment into routine use.[24] We designed analyses that assessed concurrent, convergent, predictive and content validity. The ideal for convergent validity would have been another ADL assessment. Such data were not available within VISTA, this concurs with our previous findings that BI is the most prevalent ADL assessment in trials and other measures are infrequently used.[2] We assessed agreement with similar outcome assessment scales (mRS and NIHSS) and with assessments that measure differing constructs (EQ-5D, Sheffield Test). Assessing BI against mRS and NIHSS is in keeping with previous work looking at stroke outcome properties.[25] The weaker agreement with the aphasia and quality of life scales supports the short forms as tools describing ADL rather than generic measures of stroke recovery.

We feel confident of the properties of the 3-item SF-BI that we recommend, as it performed well in our validation analyses and previous derivation and validation studies have been described. We note also that for some of our convergent validity analyses, the short forms performed better than the full BI. This may suggest that as a prognostic tool or case-mix adjuster, baseline short forms of the BI may be preferable to the full assessment.

In creating a shorter version of an existing scale there is a compromise between ease of use and richness of data captured. Standard BI is already a reasonably short assessment scale, in fact various groups have suggested that BI lacks granularity and have proposed additional items be added to the scoring or the scale. [26,27] We do not envisage the SF-BIs being used for individual clinical assessment, rather we think the short scales will have utility in large scale audit, epidemiology and clinical research. Time required for testing is a major factor in determining acceptability of a scale to therapists.[28] The SF-BIs described in the literature had a minimum of three items but assessments could be made shorter still. Our factor analysis suggests two main factors within BI, in keeping with previous descriptions. [29] There is a literature describing utility of single question assessments for certain disease states. [30]

Having suggested a promising 3-item SF-BI, the next step would be to use this short form assessment and describe if it offers any benefit over traditional BI in terms of feasibility, acceptability and completion rates. We speculate that a 3-item SF-BI will lessen assessment time, lessen test burden for patients and lead to fewer data transcription errors but all of this remains to be proven. We are encouraged that large audits, registries and clinical trials are already incorporating SF-BI into their test batteries and we would encourage any groups using the short form to share their experiences with the stroke community.

There is emerging best practice guidance on derivation and validation of short forms assessments.[31] The papers included in our review pre-date this guidance and so the variation in conduct and reporting across the studies is understandable. There is no consensus tool for "quality" assessment of such studies. We felt, as a minimum, papers should describe their derivation cohort and method; use at least two differing validation techniques and have further validation in an independent dataset. Few of the SF-BI

described in the literature fulfilled all these criteria and our VISTA based analysis assists by providing robust, multimodal validation in a large, external dataset.

The strengths of our approach include a robust literature search with internal and external validity checks and access to a large dataset of study quality data. The size of the VISTA resource allowed us to look at properties of BI with a greater precision than previously described. We recognise that our literature review included a relatively limited scope of databases, with no meta-analyses or quality assessment. The purpose was to discover SF-BI and our internal checks suggest we achieved this. A limitation of our study is around generalisability of the VISTA population. VISTA data are from randomised controlled trials and participants may not be representative of unselected stroke admissions. This is less of an issue as we propose that the SF-BI be used for audit and research purposes rather than individual patient clinical assessment. Our focus was stroke, as VISTA is a stroke specific resource and BI is often used in stroke trials. We suspect that our SF-BI could be used in non-stroke populations. However, we found few published papers describing SF-BI in non-stroke settings. Where data were available properties seemed favourable[32] but further validation work would be needed before we recommend SF-BI for other conditions. We recognise that validating a short form does not address some of the inherent limitations of the BI as measure of ADL[33], but the shortened scale should, at least, address the issue of efficiency of assessment.

Our data support use of a shortened Barthel for assessment of stroke populations. Based on multi-modal validation analyses, we recommend a three item scale that sums ability to transfer, ability to mobilise and bladder control. We hope that this short form may prove useful in future large scale trials, registries and audit.

Disclosures: None

Sources of funding: This work was supported by Greater Glasgow and Clyde Endowments, funder had no role in analysis.

Dr Quinn is supported by a Stroke Association/Chief Scientist Office Senior Clinical Lectureship.

References

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

2. Quinn TJ, Dawson J, Walters MR, Lees KR. Functional outcome measures in contemporary stroke trials.Int J Stroke.2009;4:200-205.

3. Quinn TJ, McArthur K, Ellis G, Stott DJ. Functional assessment in older people.BMJ.2011; 343:d4681.

4. Wade DT, Collin C.The Barthel ADL Index:a standard measure of physical disability?Int Dis Studies.1988;10:64–67.

5. Granger CV, Dewis LS, Peters NC, Sherwood CC, Barrett JE.Stroke rehabilitation: analysis of repeated Barthel index measures.Arch Phys Med Rehabil.1979;60:14-7.

6. Quinn TJ, Langhorne P, Stott DJ.Barthel index for stroke trials: Development, properties, and application.Stroke.2011;42:1146-1151

7. Ritchie CW, Terrera GM, Quinn TJ. Dementia trials and dementia tribulations: Methodological and analytical challenges in dementia research.Alzheimer's research & therapy.2015;7:31

8. Stone SP, Ali B, Auberleek I, Thompsell A, Young A. The Barthel index in clinical practice: use on a rehabilitation ward for elderly people.J Roy Coll Phys.1994;28:419-423.

9. Irwin P, Rutledge Z, Lowe DA. Report on the national sentinel audit of stroke. London, UK: Royal College of Physicians; 1999.

10. Kwok CS, Potter JF, Dalton G, George A, Metcalf AK, Ngeh J, et al. The soar stroke score predicts inpatient and 7-day mortality in acute stroke.Stroke.2013;44:2010-2012

11. Murray J, Forster A, Young J. Response and completion rates for postal outcomes booklets in stroke rehabilitation.Int J Ther Rehabil.2007;14:440–445

12. Hobart JC, Thompson AJ. The five item Barthel Index.J Neurol Neurosurg Psych.2001;71:225–230.

13. Ellul J, Watkins C, Barer D. Estimating total Barthel scores from just three items:the European Stroke Database “minimum dataset” for assessing functional status at discharge from hospital.Age Ageing.1998;27:115–122.

14. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses:The PRISMA Statement.BMJ.2009;339:b2535.

15. Wright JG, Feinstein AR.A comparative contrast of clinimetric and psychometric methods for constructing indices and rating scales.J Clin Epidemiol.1992;45:1201–1208.

16. MacIsaac R, Ali M, Peters M, English C, Rodgers H, Jenkinson C et al. Derivation and validation of a modified short form of the Stroke Impact Scale.Journal of the American Heart Association 2016;5:e003108.

17. Ali M, Fulton R, Quinn T, Brady M. How well do standard stroke outcome measures reflect quality of life? A retrospective analysis of clinical trial data. Stroke.2013;44:3161-3165

18. Bohannon RW. Estimating total Barthel scores from just three items.Age ageing.2004;33:321-322.

19. Bohannon RW, Landes M. Reliability, Validity, and Responsiveness of a 3-Item Barthel for Characterizing the Physical Function of Patients Hospitalized for Acute Stroke.Journal of Neurological Physical Therapy.2004;28:110-113.

20. Hsueh IP, Lin JH, Jeng JS, Hsieh CL. Comparison of the psychometric characteristics of the functional independence measure, 5 item barthel index, and 10 item barthel index in patients with stroke.J Neurol, Neurosurg Psych.2002;73:188-190.

21. Granger CV, Hamilton BB, Gresham GE. The stroke rehabilitation outcome study: part II. Relative merits of the total Barthel index and a 4-item subscore in predicting patient outcomes.Arch Phys Med Rehabil.1989;70:100–103.

22. Lekamwasam S, karunatilake K, Kankanamge SKP, Lekamwasam V. Physical dependency of elderly and physically disabled; measurement concordance between 10-item Barthel index and 5-item shorter version.Ceylon Med J.2011; 56:114-118.

23. Cho HJ, Lim YJ. Efficacy and safety of oral citicoline in acute ischaemic stroke: drug surveillance study in 4,191 cases.Experimental Clin Pharmacol.2009;31:171-176.

24. Shah S. In praise of the biometric and psychometric qualities of the Barthel Index.Physiotherapy 1994;80:769-771.

25. Uyttenboogaart M, Luijckx GJ, Vroomen PC, Stewart RE, De Keyser J. Measuring disability in stroke: relationship between the modified Rankin scale and the Barthel index.J Neurol.2007; 254:1113-1117.

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

27. Fortinsky RH, Granger CV, Seltzer GB. The use of functional assessments in understanding home care needs. Med Care. 1981;19:489–497.

28. McAvoy E. The use of ADL indices by occupational therapists.Brit J Occup Therapy.1991;54:383-385.

29. de Morton NA, Keating JL, Davidson M. Rasch analysis of the Barthel Index in the assessment of Hospitalized older patients after admission for an acute medical condition.Arch Phys Med Rehabil.20008;89:641-647.

30. Hendry K, Hill E, Quinn TJ, Evans J, Stott DJ.Single screening question for cognitive impairment in older people: a systematic review.Age Ageing.2015;44:322-326.

31. Goetz C, Coste J, Lemetayer F, Rat AC, Montel S, Recchia S, et al. Item reduction based on rigorous methodological guidelines is necessary to maintain validity when shortening composite measurement scales. J Clin Epidemiol.2013;66:710-718.

32. Tagharrobi Z, Sharifi K, Sooky Z, Tagharrobi L. Abstract: Psychometric evaluation of the Barthel Index short forms in Iranian elderly people residing in nursing homes. Presented at "Controversies in Longevity, Health and Aging (CoLONGY)".Barcelona, 2010.

33. Harrison JK, McArthur KS, Quinn TJ. Assessment scales in stroke:clinimetric and clinical considerations.Clin Interv Ageing. 2013;8:201-211.

Table 1. Papers describing short forms of the Barthel Index

	Populations assessed	**Derivation cohort "n"**	**Item reduction method**	**Validation method**	**Tested in external dataset**
Bohannon 3-item (two papers)[14,15]	Stroke unit admissions	251, 275	N/A	Multiple validation analyses	Validation of Ellul scale with modifications
Cho[19]	Acute ischaemic stroke	N/A	N/A	N/A	N/A
Ellul 3-item[9]	Stroke unit admissions	169	Predictive ability of combinations of items	Unclear	Stroke RCT
Granger 4-item[17]	Stroke Rehabilitation Outcome Study	539	Four items associated with independence	Predictive (discriminant) validity	No
Hobart 5,4 and 3-item[8]	Neurological rehabilitation unit	844	Corrected item total correlation, effect size	Multiple validation analyses	Dataset split 50:50 derivation validation
Hseuh 5-item[16]	Stroke Unit admissions	125	N/A	Multiple validation analyses	Validation of the Hobart scale
Lekamwasam 5-item[18]	Medical and orthopaedic clinics	286	Factor analysis	Correlation with standard BI	Validation of Hobart scale (modified)

Table 2: Correlation (Spearman's rho) for each Barthel Index item with total Barthel Index score

Barthel Index attribute	**Day 90 Acute data**
1.Feeding	0.82
2.Bathing	0.84
3.Grooming	0.72
4.Dressing	0.89
5.Bowel Control	0.66
6.Bladder Control	0.70
7.Toilet Use	0.85
8.Transfer	0.85
9.Mobility	0.85
10.Stairs	0.88

Table 3: Correlation of various forms of Barthel Index with other outcomes

Scale	BI full scale	Cho 3-item	Ellul 3-item	Granger 4-item	Hobart 5-item	Hobart 4-item	Hobart 3-item	VISTA 3-item
BI90 full scale	1.00	0.95	0.90	0.88	0.95	0.94	0.92	0.90
mRS90	-0.90	-0.89	-0.85	-0.81	-0.90	-0.89	-0.87	-0.85
NIHSS 90	-0.81	-0.81	-0.74	-0.74	-0.79	-0.79	-0.78	-0.75
EQ-5D M3	0.53	0.48	0.47	0.42	0.49	0.48	0.48	0.47
EQ-5D VAS M3	0.24	0.23	0.19	0.17	0.21	0.23	0.21	0.19
SIS full scale	0.62	0.60	0.52	0.48	0.57	0.58	0.55	0.53
SSTALD	0.31	0.28	0.28	0.36	0.27	0.28	0.27	0.28
Cho 3-item	0.95	1.00	0.87	0.88	0.94	0.94	0.91	0.87
Ellul 3-item	0.89	0.87	1.00	0.85	0.92	0.90	0.90	0.99
Granger 4-item	0.88	0.88	0.85	1.00	0.83	0.83	0.84	0.85
Hobart 5-item	0.95	0.94	0.92	0.83	1.00	0.99	0.97	0.92
Hobart 4-item	0.94	0.94	0.90	0.83	0.99	1.00	0.97	0.91
Hobart 3-item	0.92	0.91	0.90	0.84	0.97	0.97	1.00	0.90
VISTA 3-Item	0.90	0.87	1.00	0.85	0.92	0.91	0.90	1.00

Correlation coefficient (rho) between each outcome measure and SF-BI. For non-standardised variables spearman rank correlation coefficient was used. All significant at pre-specified level ($p<0.001$).

BI=Barthel Index; SF-BI=short form Barthel Index; mRS=modified Rankin Scale; NIHSS=National Institutes of Health Stroke Scale; EQ-5D=Euro-Qol 5 dimension; SIS=Stroke Impact Scale; SSTALD=Sheffield Screening Test for Acquired Language Disorder

Table 4: Association of various forms of Barthel Index at baseline with 90 day outcomes

Variable	mRS			NIHSS		
	OR (95%CI)	**R-Sq**	**C stat**	**OR (95%CI)**	**R-Sq**	**C stat**
BI	1.24 (1.23,1.25)	0.84	0.91	1.08 (1.08,1.085)	0.69	0.83
Cho 3-item	1.52 (1.49, 1.55)	0.80	0.90	1.26 (1.25,1.28)	0.68	0.83
Ellul 3-item	1.19 (1.18,1.20)	0.76	0.87	1.07 (1.07,1.07)	0.62	0.80
Granger 4-item	1.35 (1.33,1.38)	0.85	0.67	1.20 (1.19,1.21)	0.59	0.80
Hobart 5-item	1.39 (1.36,1.41)	0.82	0.90	1.13 (1.12,1.13)	0.67	0.82
Hobart 4-item	1.44 (1.41,1.47)	0.81	0.90	1.18 (1.170,1.19)	0.67	0.82
Hobart 3-item	1.55 (1.51,1.58)	0.78	0.88	1.24 (1.23,1.25)	0.65	0.82
VISTA 3-item	1.50 (1.47,1.54)	0.76	0.87	1.18 (1.17,1.18)	0.62	0.80

Multivariate ordinal regressions showing the relationships between each SF-BI and outcome measures. Values are odds of better outcome on mRS and NIHSS, adjusted for age,sex,stroke type.