The measurement properties of the original Barthel Index
and its applicability to measure function with older adults:
A systematic review

Elizabeth Wyatt

Supervisor:
Kylie Wales

Thesis presented in partial fulfilment of the degree of
Masters of Occupational Therapy

Discipline of Occupational Therapy
Faculty of Health Sciences
The University of Sydney
2013
STATEMENT OF AUTHENTIFICATION

I, Elizabeth Wyatt, hereby declare that this submission is my own work and that it contains no material previously published or written by another person. Nor does it contain any material that has been previously accepted for another degree.

Name: Elizabeth Wyatt
Signed: _____________
Dated: _____________
ACKNOWLEDGEMENTS

I would personally like to thank my supervisor Kylie Wales who provided her time, wisdom, guidance, support and encouragement in the completion of this thesis.

I would also like to personally thank Lynette MacKenzie, who without her commitment to research and the Masters of Occupational Therapy program at the University of Sydney the research elective embedded within this degree would not have been available for Masters Students.

To my fellow Honours students Charmain, Emma and Nicole I appreciate your support, encouragement and laughs during the year. This has been a wonderful experience for me to share this journey with likeminded individuals who are also passionate about research.
Background: Worldwide many countries are experiencing ageing populations. Ageing can result in older adults experiencing compromised health and increased hospitalisations making older adults recurrent users of health care services. The process of ageing can reduce an older adult’s ability to perform everyday tasks and participate in the activities required for daily life. As a result, older adults are often referred to occupational therapists to improve functional capacity and support engagement in activities and occupations. One role occupational therapists perform when working with older adults in a hospital or rehabilitation setting is to enhance function in self-care tasks and plan for discharge. Occupational therapists use functional assessments to determine an older adult’s ability to carry out everyday tasks. Some occupational therapists use standardised assessments whereas others prefer to use non-standardised assessments. The major barriers to occupational therapists using non-standardised assessments are a lack of knowledge and skill. Standardised assessments are used to measure function, monitor change and determine the effectiveness of therapy. The quality of these measurements relies on the validity and reliability of standardised assessments when used with specific patient groups and clinical settings they were designed for. The original Barthel Index (BI) developed by Mahoney and Barthel (1965) is a standardised functional assessment widely used within rehabilitation, hospital and community settings with various patient groups. To date, the validity and reliability of the original BI has not been established with older adults.
**Aim:** The aim of this research is to investigate the measurement properties of the original BI and to determine its suitability as a standardised assessment to measure function with older adults.

**Method:** A comprehensive systematic review of the literature was undertaken and the following databases were searched; Cumulative Index of Nursing and Allied Health (CINAHL), Embase and Medline. Methodological quality of the included studies was assessed using the **CO**nsensus-based **St**andard for the selection of health **M**easurement **IN**struments (COSMIN) checklist.

**Results:** 3273 abstracts were screened and 5 articles included. COSMIN rated structural validity (good), cross-cultural validity (poor), hypothesis testing (fair to poor), internal consistency (poor) and responsiveness (fair). No studies investigated reliability, measurement error, criterion or content validity.

**Discussion:** Limited research has been undertaken on the measurement properties of the original BI, highlighting the need for further robust research to be undertaken on the measurement properties of the original BI, so that the reliability, validity and limitations of the original BI to be established with older adults. Multiple versions of the BI were identified making interpretation of results problematic. This review recommends re-naming these versions to ensure transparency and straightforward interpretation of results.
**Implications:**

This and further research will continue to bridge knowledge gaps within the occupational therapy profession and enable occupational therapists to make informed decision when selecting the most appropriate standardised assessment to measure function with older adults. Breaking down the barriers to measuring outcomes through increased knowledge and use of standardised assessments will increase occupational therapists adherence to evidence base practice and professional accountability to the occupational therapy profession. This in turn will provide older adults with accurate measurements of their functional capacity which will ultimately impact on patient recovery, quality of care and earlier discharge.
TABLE OF CONTENTS

Title page………………………………………………………………………………. 2
Statement of authentication……………………………………………………………. 3
Acknowledgements…………………………………………………………………….. 4
Thesis Abstract………………………………………………………………………… 5
Table of Contents………………………………………………………………………. 8
List of Tables……………………………………………………………... 11
List of Figures………………………………………………………………………….. 12
List of Appendices……………………………………………………………………... 13

SECTION 1: LITERATURE REVIEW

1. Introduction
   1.1 Background to the topic…………………………………………………………. 15
   1.2 Aim of the literature review…………………………………………………… 17
   1.3 Search strategy…………………………………………………………………… 18
   1.4 Theoretical Framework………………………………………………………… 18

2. Measurement properties…………………………………………………………. 19

3. The COnsensus-based Standards for the selection of health Measurement
   Instruments (COSMIN) checklist ………………………………………………… 19

4. Measurement………………………………………………………………………. 21

5. Standardised assessments……………………………………………………….. 23

6. Measuring outcomes in occupational therapy………………………………… 23
7. Barriers to measuring outcomes in occupational therapy .......................... 24

8. Overview of the literature on the Barthel Index
   8.1 The original Barthel Index ....................................................... 25
   8.2 Versions of the Barthel Index ................................................... 26
   8.3 Application of the Barthel Index to clinical practice ......................... 27
   8.4 Previous research on the Barthel Index ....................................... 28
   8.5 Previous research on the original Barthel Index ................................ 30
   8.6 Application of the original Barthel Index to older adults .................. 32

9. Issues arising from the literature ..................................................... 32

10. Statement of research problem ....................................................... 33
    10.1 Research aims .................................................................. 33
    10.2 Research questions ............................................................. 34

11. Conclusion .................................................................................... 34

12. References .................................................................................... 35

SECTION 2: JOURNAL MANUSCRIPT

Title page ......................................................................................... 47

Abstract ......................................................................................... 48

Introduction ..................................................................................... 49

The original Barthel Index ............................................................... 51

Measurement properties of the original Barthel Index ...................... 52

Methods .......................................................................................... 53

Search strategy ................................................................................ 53

Screening ......................................................................................... 53
Evaluation of included studies ......................................................... 54
COSMIN rules .................................................................................. 56
Results .............................................................................................. 56
Structural validity .............................................................................. 58
Cross cultural validity ....................................................................... 59
Hypothesis testing ............................................................................ 60
Internal consistency ........................................................................... 61
Responsiveness .................................................................................. 62
Application to older adults ................................................................. 63
Discussion ......................................................................................... 64
Limitations ......................................................................................... 67
Implications ......................................................................................... 68
Recommendations ............................................................................... 68
Conclusions ......................................................................................... 69
References .......................................................................................... 71
LIST OF TABLES

SECTION 2

Table 1: COSMIN checklist definitions of measurement properties .......... 78
Table 2: Characteristics of included studies .................................... 79
Table 3: Reasons for exclusion at secondary screening ......................... 81
Table 4: Methodological quality of the included studies ...................... 82
LIST OF FIGURES

SECTION 2

Figure 1: Screening process ................................................................. 83
LIST OF APPENDICES

SECTION 2

Appendix 1:  Search strategy................................................................. 84
Appendix 2:  Reasons for paper exclusion at secondary screening............. 90
Appendix 3:  Physical & Occupational Therapy in Geriatrics Author Guidelines... 93
SECTION 1

REVIEW OF THE LITERATURE

The measurement properties of the original Barthel Index and its applicability to measure function with older adults

Elizabeth Wyatt
The measurement properties of the original Barthel Index and its applicability to measure function with older adults

1. Introduction

1.1 Background to the topic

Throughout the world many populations are seeing an increasing number of older adults surviving to very old age (Crews & Zavotka, 2006). In countries such as Australia, Japan, Norway, Sweden, the United States of America and the United Kingdom the number of older adults per capita has increased up to 15% of the total population (Crews & Zavotka, 2006; Cheah & Presnell, 2011). The consequences of ageing include increased frailty and incidence of falls, functional and cognitive decline and increased levels of disability or disease (Crews & Zavotka, 2006). Many of these conditions result in hospitalisation or rehabilitation, making older adults recurrent users of health care services (Giles, Cameron, & Crotty, 2003).

Ageing can result in the reduced ability of older adults to perform everyday tasks such as activities of daily living (ADL’s) and instrumental activities of daily living (IADL’s) (Shearer & Guthrie, 2013; Wallace et al., 2002). ADL’s are the self-care tasks individuals need to perform and include feeding, toileting, bathing, grooming and mobilisation (Shearer & Guthrie, 2013). IADL’s are activities that support daily life within the home or community (Wallace et al., 2002). Many older adults strongly value their independence and ability to complete everyday tasks and participate in social activities, all of which contribute to the quality of their daily lives (Wallace et al., 2002). When older adults are experiencing the
effects of ageing they are often referred to occupational therapists to improve functional
capacity and support re-engagement in the activities they find meaningful.

The profession of occupational therapy is concerned with occupational engagement,
promoting well-being and quality of life (Tomaszewski, 2013). The term occupation
describes everything people do in their lives that provides a sense of meaning and fulfilment
(Law, Baum, & Dunn, 2001). This includes independently performing tasks and participating
in social and leisure based activities (Law, Steinwender, & Leclair, 1998). Occupational
therapists work with older adults in acute care or rehabilitation settings to increase functional
capacity in self-care tasks and plan for discharge (Wales, Clemson, Lannin, & Cameron,
2012; Barras, 2005; Parks & Shepperd, 2000; Sheperd et al., 2013). Occupational
therapists achieve this by using standardised assessments to measure outcomes and the
effectiveness of treatment. (Unsworth, 2000).

Barriers to measuring outcomes exist in clinical practice (Bowman, 2006). The literature
suggests that some therapists use non-standardised assessments to measure outcomes whereas
others use standardised assessments (Unsworth, 2000; Bowman, 2006; Muller et al., 2011,
Fawcett, 2007). Non-standardised assessments have not been meticulously studied or
scientifically explored resulting in an inability to quantify outcomes or verify treatment
effectiveness (Unsworth, 2000; Law & Letts, 1989; Fawcett 2007). Standardised assessments
have been developed through rigorous scientific process enabling therapists to objectively
measure patient change, investigate the effects of therapeutic interventions and are used for
benchmarking (Fawcett, 2007; Unsworth, 2000; Law & Letts, 1989; Wales et al., 2012).
Many types of standardised assessments designed to measure function exist, however this review will concentrate on one standardised assessment, the original Barthel Index (BI).

Many versions of the BI exist in the literature leading to confusion surrounding nomenclature. Therefore this review will adopt the terminology of ‘BI’ when referring to all versions of the Barthel Index and the term ‘original BI’ when discussing the original version of the Barthel Index developed by Mahoney and Barthel. 1965.

The original BI was developed by Mahoney & Barthel in 1965 to measure functional change in an orthopaedic in-patient rehabilitation setting, prior, during and at the end of treatment. The literature suggests the BI has been used widely in a variety of patient groups and environmental settings (Fricke & Unsworth, 1996; Wade, 1987). The BI is commonly used with older adults particularly in stroke rehabilitation and acute care settings however establishing the validity and reliability of the original BI and its ability to measure function with older adults has not been determined (Kwon, Hartzema, Duncan, & Min-Lai, 2004; Quinn, Langhorn, & Scott, 2011; Boger, Demain, & Latter, 2012).

1.2 Aim of the literature review

The literature review will report on the barriers occupational therapists face using standardised assessments and how this impacts their ability to measure outcomes, monitor patient change and treatment effectiveness. The review will also examine what is currently known about the measurement properties of the original BI and whether the original BI is a valid and reliable assessment tool that measures function with older adults.
1.3 Search strategy

Literature for this review was sourced via a comprehensive search of the following electronic databases; Cumulative Index of Nursing and Allied Health (CINAHL), Medline (via Ovid SP), Occupational Therapy Systematic Evaluation of Evidence (OTseeker), the Cochrane Library and Web of Science. Search terms included Occupational therap*, Barthel Index, Adult*, Age*, Geriatric*, Acute Care, Hospitalisation, Community Health Service*, Hospital*, Rehabilitation, Patient discharge, Outcome assessment (Health Care), Valid*, Reliab* Psychometric* and Reproducibility of results. Different combinations of search terms were applied to provide a wide literature source. Articles were screened at abstract and relevant articles retained. Reference lists from included articles were reviewed to determine if any additional publications would add more information to the review. These articles were sourced via the University’s online library system. There was no predetermined time limitations as all articles from the earliest available to date were deemed relevant. Studies were limited to those published in English.

1.4 Theoretical Framework

Theoretical frameworks provide a model to which the fundamental concepts of the profession guide occupational therapists in clinical practice (Rogers, 2005). The International Classification of Functioning, Disability and Health [ICF], (2001), provides a framework to classify the contexts of human functioning and disability, activity and participation. Activity and participation is incorporated under function in the ICF model which aligns with the client centred approach inherent to the occupational therapy profession (Perenboom & Chorus, 2003; Muller et al., 2011; Glassel et al., 2010; Arnadottir, Gunnarsdottir, Stenlund,
Lumdin-Olsen, 2011). Impairment is also considered under function, however, occupational therapists tend to work within the activity and participation framework (Wales et al., 2012). Since the profession of occupational therapy is concerned with these interrelationships the ICF will be used as the framework for this review and throughout this paper the term function will be used to encompass activity and participation.

2. Measurement Properties

The term measurement properties are an overarching term used to define reliability and validity and may be used interchangeably with titles such as psychometric properties (Mokkink et al., 2010). Measurement properties have varying definitions throughout the literature leading to inconsistencies in evaluation and reporting (Mokkink et al., 2010). Identifying uniformity in relation to the terms and definitions for each measurement property is essential as it aids in consistent global reporting of experimental design, process and the statistical outcomes investigated for future research (Mokkink et al., 2010). This review acknowledges disagreement in the literature regarding the terminology surrounding measurement properties and will therefore adopt the terminology outlined on page 20-21 throughout this review for consistency.

3. The Consensus-based Standard for the selection of health Measurement Instruments (COSMIN) checklist

This review acknowledges alternate methods to assess measurement properties. The Consensus-based Standard for the selection of health Measurement Instruments (COSMIN) checklist was selected for this research project to evaluate the measurement properties of the original BI. The reasons the COSMIN checklist was selected include that an international
Delphi review was conducted using 43 international experts across the fields of psychology, epidemiology, clinical medicine and statistics. Multiple surveys were conducted before consensus was reached to determine which measurement properties would be evaluated and how they would be defined. This process led to the development of an international standard ensuring global consistency in relation to how measurement properties are evaluated in terms of study design and statistical evaluation (de Vet et al., 2013). This process achieved agreement in the definitions and domains of measurement properties, ensuring consistent terminology can be used across contexts with uniformity (Mokkink et al., 2010). The COSMIN checklist defines measurement properties as summarised in Table 1.

Table 1: COSMIN checklist definitions of measurement properties (Mokkink et al., 2010)

<table>
<thead>
<tr>
<th>Measurement Property</th>
<th>COSMIN checklist definitions of Measurement Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>Is associated with the interconnectedness of the assessment items. Internal consistency crosses into the domain of reliability in relation to measurement error</td>
</tr>
<tr>
<td>Reliability</td>
<td>Is concerned with the degree measurement is free from measurement error when the measure is repeated under several conditions with unchanging patients. The domain of reliability consists of test-retest reliability, inter-rater reliability (by different raters on the same occasion) and intra-rater reliability (by the same rater on different occasions)</td>
</tr>
<tr>
<td>Measurement error</td>
<td>Investigates random and systematic errors of patient scores</td>
</tr>
<tr>
<td>Content validity</td>
<td>Explores the degree to which the content of an assessment tool measures the construct it is measuring</td>
</tr>
<tr>
<td>Structural validity</td>
<td>Is concerned with the degree to which the measurement of an assessment tool accurately reflects the construct being measured</td>
</tr>
<tr>
<td>Hypothesis testing</td>
<td>Investigates the relationships between the construct being measured and the assessment tool. Hypothesis testing can be</td>
</tr>
</tbody>
</table>
The COSMIN checklist classifies measurement properties into nine domains. They include; internal consistency, reliability, measurement error, content and criterion validity, structural validity, hypothesis testing, cross-cultural validity and responsiveness. Each study can be assessed under more than one domain. Using predefined criteria the measurement properties of each study can be evaluated. The COSMIN checklist rates the measurement properties as excellent, good, fair or poor. To ensure the methodological quality of the entire study, the overall rating for methodological quality equals the lowest rating received per measurement property (de Vet et al., 2013). Therefore the proposed research will use the COSMIN checklist as the tool to assess the methodological quality of the measurement properties of the included studies.

4. **Measurement**

Measurement can be used for the purpose of evaluation, informing decisions, understanding relationships between concepts and constructs, demonstrating change, predicting outcomes and drawing conclusions (Portney & Watkins, 2009). Establishing whether a standardised
assessment can produce valid and reliable result is essential for use in clinical practice as the precision of the measurements validate the effectiveness of therapeutic interventions. This in turn adds strength to evidence based practice and benchmarking practices of the occupational therapy profession (O’Connor, Cano, Thompson, & Hobart, 2004; Portney & Watkins, 2009).

In clinical practice, occupational therapists use standardised assessments to evaluate the patient’s current medical status, monitor progress, prescribe treatment, determine the effectiveness of treatment, and predict patient outcomes and plan for discharge (Unsworth, 2000). It is imperative that occupational therapists understand the importance of using standardised assessments with demonstrated validity and reliability within target populations and environmental settings (Unsworth, 2000). Otherwise therapists may measure outcomes inaccurately which can have detrimental effects on patient recovery and delay discharge resulting in increased costs to the healthcare system.

Reliability is concerned with the extent to which a repeated measurement yields the same result. Validity is concerned with the extent to which an instrument measures what it is intended to measure (Portney & Watkins, 2009; Gosman-Hedstrom & Svensson, 2000). Ensuring the reliability and validity of standardised assessments when used with specific patient groups and clinical settings is essential. Determining the reliability and validity of the standardised assessment ultimately ensures correct measurements and positive patient outcomes. This can be achieved by demonstrating the value of therapy in improved patient recover and faster discharge home.
5. **Standardised assessments**

Standardised assessments are tools that have been developed through a rigorous and scientific manner for a defined construct and population group (Fawcett, 2007). Standardised assessments assign a mathematic value of quantifiable means to a construct which then allows comparison between individuals (Law, 1987). The process of administration, guidelines for delivering the assessment and the scoring criteria are clearly defined to allow objective measurement of patient change and the effectiveness of treatment (Fawcett, 2007; Unsworth, 2000; Law & Letts, 1989). Standardised assessments require proven reliability and validity to be used successfully with different population groups, clinical settings and medical conditions other than those they were originally intended for (Unsworth, 2000).

6. **Measuring outcomes in occupational therapy**

Measuring outcomes using standardised assessments provides information related to client progression while establishing the value and effectiveness of treatment (Unsworth, 2000). Measuring outcomes enables occupational therapists to identify a patient’s current functional status and using this information subsequently direct therapy (Duncan & Murray, 2012; Hinkle, Davies, Ng, & McClaren, 2008). Selecting an appropriate standardised assessment can be time consuming as no standardised assessment will measure all the domains for every client group or environmental setting and the advantages and disadvantages of each assessment must be considered (Fricke & Unsworth, 1996; Kjeken, 2012). Therefore it is vital that occupational therapists select an appropriate standardised assessment to assess function that has demonstrated validity and reliability with older adults (Kjeken, 2012; Fricke & Unsworth, 1996).
In acute care and rehabilitation settings occupational therapists use standardised assessments to measure older adult functional capacity and predict recovery for discharge (Law, 1987; Duncan & Murray, 2012; Barras, 2005). Discharge planning encompasses the development of an individualised discharge plan for each patient before leaving hospital with the aim to contain costs and improve patient outcomes (Parks & Shepperd, 2000). Evidence suggests that discharge planning reduces patient time in hospital and can prevent readmission (Shepperd et al., 2013).

7. **Barriers to measuring outcomes in occupational therapy**

Occupational therapists report that measuring outcomes is challenging in clinical practice because they have difficulty identifying what to measure, how to classify function as a measureable unit and difficulty linking these to measureable goals (Bowman, 2006). Furthermore, a lack of knowledge and skill, time, budgetary restraints and lack of managerial support have been cited as the possible reasons for inconsistent use (Law, 1987; Bowman, 2006; Atwal, McIntyre, Craik, & Hunt, 2007; Bowman & Llewellyn, 2002; Law & Letts, 1989). Other barriers include a lack of individual understanding of the importance reliability and validity play when using standardised assessments to measure outcomes and treatment effectiveness (Fawcett, 2007; Bowman, 2006). Alternatively, the reasons may be more personal in nature and may conflict with the values or beliefs of the individual occupational therapist (Bowman, 2006; Atwal et al., 2007; Bowman & Llewellyn, 2002; Law & Letts, 1989). As a result of such difficulties, some occupational therapists tend to favour the use of non-standardised assessments that can easily be adapted to suit the client needs and clinical setting in which the occupational therapists work (Bowman, 2006; Atwal et al., 2007; Law & Letts, 1989). However, the discrepancies in non-standardised assessments can result in an
inability to quantify outcomes or verify treatment effectiveness which can significantly impact on the patient recovery, cost of treatment and length of stay (Unsworth, 2000; Law & Letts, 1989).

In the focus group study conducted by Bowman (2006) occupational therapists acknowledged they experienced difficulty in measurement but understood that demonstrating clinical effectiveness was one way the occupational therapy profession could demonstrate professional credibility. The group acknowledged the profession needs to change the way outcomes are measured and commence using evidence based principles (Bowman, 2006). Occupational therapists recognise the importance of using standardised assessments to effectively measure outcomes and determine treatment effectiveness, however lack of knowledge surrounding the selection an implementation of standardised assessments requires education, training and industry support (Bowman, 2006; Kjeken, 2012).

8. Overview of the literature on the Barthel Index

8.1 The original Barthel Index

Mahoney and Barthel (1965) developed an assessment tool to evaluate functional improvement over time of orthopaedic patients in an in-patient rehabilitation setting, prior, during and at the end of rehabilitation (Mahoney & Barthel, 1965; Fricke & Unsworth, 1996). The original BI measures 10 areas of ADL related to functional independence or dependence in self-care tasks and mobility. These include wheelchair-bed transfer, personal toilet, toilet transfer, bathing, walking, stair climbing, dressing and bladder and bowel control. The original BI uses an ordinal scale scoring system ranging from 0 to 100 points in 5-point increments. A score closer to 100 indicates increased functionality whereas scores closer to 0
represent increased dependency (Mahoney & Barthel, 1965). The literature suggests the BI is commonly used with older adults however the validity and reliability of the original BI and its ability to measure function with older adults has not been established to date (Fricke & Unsworth, 1996).

8.2 Versions of the Barthel Index

The literature review uncovered many versions of the BI. A commonly used version was developed by Collin & Wade (1988) who adapted the scoring and guidelines. This version continues to be referred to as the ‘Barthel Index’ which adds confusion to the application of the original BI. Inexperienced readers may misinterpret the results of a study, believing they were researching the original BI when in fact the results of the study utilise the Collin and Wade scoring which may produce different results to the original BI. Some examples of studies where different scoring was used but the title was misleading where in the study by Hsueh and colleagues (2001) who investigated the measurement properties of the BI in stroke, the study by Frick & Unsworth (1996) who investigated inter-rater reliability in adult rehabilitation and in the study by Hinkle and colleagues (2008) who examined various assessment tools for discharge planning in an acute care setting. All these studies use the term ‘the Barthel index’ however on closer investigation the Collin and Wade scoring was used so these results cannot be generalised to the original BI.

Other versions of the BI include variations in the delivery method (by self-report, via telephone, postal survey or proxy), in the length as a shortened (3-item and 5-item version) or extended version, or through the use of translations (Oveisgharan et al., 2006; Hobart, Cano, & Thompson, 2010; Loar, Haig, Yamakawa, & Baljinnyam, 2011, Bohannon, & Landes,
2004; Hsueh, Lin, Jeng, & Hsieh, 2002; Jansa, Pogacnik, & Gompertz, 2004). As any adaption or modification made to the scoring, guidelines or delivery of a standardised assessment makes it a new assessment the above changes have resulted in many versions of the BI becoming available in the literature. All versions need be viewed independently and the measurement properties of each version need to be assessed to ensure they are valid and reliable standardised assessments suitable for use in clinical practice.

Another commonly used adaption which is easily identifiable in the literature is the version by Shah, Vanclay and Cooper (1989). Shah and colleagues adapted the original BI to have increased sensitivity to change and named this version the Modified Barthel Index (MBI). The MBI utilises adapted scoring and guidelines and as a result of the name change this version is clearly distinguishable in the literature and interpretation of the results should be evident to the reader.

**8.3 Application of the Barthel Index to Clinical Practice**

Since its inception, the BI has been used by allied health professionals in numerous patient groups, in acute care, rehabilitation and community settings delivered on admission and prior discharge to measure outcomes, inform clinical decisions, and measure treatment effectiveness (Mahoney & Barthel, 1965; Atwal et al., 2008, Wales et al., 2012). The measurement properties of the BI have been investigated in stroke rehabilitation, traumatic brain injury, and multiple sclerosis patient groups, however when reviewing which versions of the BI were used a high proportion use adapted versions (Turner-Stokes, 2002; Van Baalen et al., 2006; Hobart et al., 2001). These examples used the Collin and Wade version, Modified Barthel Index and a translated version.
8.4 Previous research on the Barthel Index

The systematic review and meta-analysis undertaken by Duffy and colleagues (2013) examined inter-rater reliability of the BI for standard administration post stroke. While the review found that the BI possessed excellent inter-rater reliability it used a mixture of versions of the BI to come to these conclusions. These included the original BI, the Collin and Wade version, a variety of translated and adapted versions and one study used the Modified Barthel Index. Three quarters of the included studies were performed in non-English speaking countries and used a translated version. It is well recognised that if standardised assessments are used across cultures the items must be translated well linguistically and also be culturally adapted so that the international version maintains content validity across cultures (Beaton et al., 2000). As these studies did not mention the quality of the translation process, any findings on the reliability and validity of the international versions cannot be assumed to be an accurate reflection of the original BI.

While the majority of the included studies in the review by Duffy and colleagues (2013) assessed stroke patients face to face, the remaining studies assessed patients via telephone, postal questionnaire and video. These adaptations do not adhere to the original delivery guidelines and therefore make these versions new assessments. The validity and reliability of these versions need to be independently assessed to determine their ability to measure function. At present, we are unable to generalise these inter-rater results to the original BI and its ability to measure function with older adults. Duffy and colleagues (2013) identified a limitation relating to the varying methodological quality of the included studies. Only two
studies were deemed to be of high methodological quality raising questions as to the trustworthiness of the majority of the results.

Sainsbury and colleagues (2005) undertook a systematic review to investigate the reliability of the BI with older adults with mixed medical conditions (excluding stroke). The authors identified a small number of studies that met the inclusion criteria and commented on the notable gaps in the literature surrounding test-retest reliability of the BI with this population group. The majority of studies included in the systematic review by Sainsbury et al (2005) utilised adapted versions, the most common being the Collin and Wade version. The results found inter-rater reliability displayed high intraclass correlation coefficients of the total scores independent of how the BI was administered, for example; via postal questionnaire, telephone or face-to-face interview. These adaptations mean the results are not truly representative of the original BI and therefore information on the reliability of the original BI cannot be drawn from the results of this study.

The systematic review undertaken by Balu (2009) compared the responsiveness of the BI and the Modified Rankin Scale (MRS) three months post stroke in a rehabilitation setting. Responsiveness is concerned with the ability of a standardised assessment to measure patient change over time (Mokkink et al., 2010). The included studies were a variety of the original BI, Collin and Wade and translated versions and therefore the results cannot solely be reflective of the original BI ability to measure responsiveness.

A common theme appears to be emerging across the literature where the term ‘Barthel Index’ is representative of all versions of the BI such as the original BI, the Collin and Wade, postal
or telephone questionnaire, self-report, shortened and extended or translated versions. When systematic reviews merge these versions together they are not comparing like with like, but instead are actually comparing different assessments making the quality of previous research on the BI questionable. These inconsistencies in nomenclature make it very difficult for researchers, occupational therapists and students to accurately interpret the results. This may result in inexperienced readers misinterpreting the true results of the studies which may impact on the therapeutic interventions in clinical practice. Therefore the proposed research will only concentrate on the original English version of the original BI to ensure results are definitive.

8.5 Previous research on the original Barthel Index

The critical review undertaken by Law and Letts (1989) evaluated 13 functional assessments against a set of methodological criteria for the purpose of informing allied health professionals as to which standardised assessment would be the most appropriate for use in their practice setting. The authors rated the original BI as having ‘excellent’ validity (content, construct and responsiveness) and ‘good’ reliability (internal consistency, inter-rater and intra-rater and test-retest). The authors also concluded that the original BI had the best ability to predict ADL function and measure change in adult rehabilitation. However the age ranges of adults were not disclosed therefore the results from this research cannot be generalised to the older adult population. The limitations of this research was that it was not a systematic review and the measurement properties were evaluated against the Standard Measure Review Form developed by Law in 1987, which is not an internationally agreed scale therefore, global consistency of the results cannot be guaranteed.
The study by Granger et al (1988) established the ability of the original BI to discriminate between patient groups (construct validity) and predict outcomes (predictive validity) through patient scores. 70% of patients who scored low on the original BI (scores ranged from 0-40) had died or were living in long term rehabilitation care post stroke, whereas, 94% of patients with high scores (ranging from 80-100) had been discharged home. These patients had higher functional ability therefore participated more socially within their local communities and ultimately experienced a higher quality of life (Granger, 1988). This research displays the ability of the original BI to predict outcomes in stroke patients. If the proposed research can establish the ability of the original BI to predict outcomes with older adults the profession of occupational therapy will be able to use the original BI with confidence when planning the discharge of older adults.

Dromerick and colleagues (2003) reported a weakness of the original BI to be the presence of floor and ceiling effects. Ceiling effects occur when the maximum score of a standardised assessment is achieved by many disabled patients, meaning that the original BI is not able to discriminate levels of disability among highly functioning patients (Eakin, 1993). Dromerick and colleagues (2003) compared the sensitivity to change in disability of four standardised assessments used with stroke patients (original BI, Expanded MRS, Functional Independence Measure and International Stroke Trial Measure). The ceiling effect of the OBI was evident with 27% of patients achieving a high score of 95-100. None of these patients achieved a high score on the other standardised assessments indicating the original BI may only measure function to a maximum level in stroke populations. These findings have been supported in the literature in a number of studies. Kwon et al (2004) and Quinn, Langhorn, and Scott
(2011) results align with Dromerick et al (2003) and indicated that the original BI may not be sensitive enough to measure higher levels of functional ability in stroke populations.

Similarly the original BI experiences floor effects when the assessment is performed too early post stroke or injury and patients are bedridden. This makes the scoring appear lower than the patients’ true functional capacity (Adams et al., 2003). Floor and ceiling effects of the original BI may result in patient scores that are not truly representative of the patients’ functional status (Dromerick et al., 2003). As a significant number of older adults experience stroke, it will be interesting to see if floor and ceiling effects are present when the original BI is used with older adults.

8.6 Application of the original Barthel Index to older adults
Limited evidence exists in the literature about the ability of the original BI to measure function with older adults. Previous research undertaken on the measurement properties of the original BI has focused on stroke populations or older adult medical patients in the studies by Wallace et al (2002), Kwon et al (2004) and de Morton et al (2008) respectively. To date we are unable to generalise these results to the greater older adult population.

9. Issues arising from the literature review
While the BI has been reported to be used in various clinical settings by allied health professionals with different population groups worldwide, there is limited evidence in the literature regarding the measurement properties of the original BI and less is known about its ability to measure function with older adults (Wales et al., 2012; Law & Letts, 1989; Mahoney & Barthel, 1965; Fricke & Unsworth, 1996; Wade, 1987). Researching the
measurement properties of the original BI is necessary to fill the knowledge gap within the occupational therapy profession in regards to the selection of an appropriate standardised assessment to measure function with older adults. Occupational therapists need to determine whether the original BI has proven validity and reliability to measure function with older adults. With this knowledge occupational therapists will be able to effectively identify whether the original BI is an appropriate standardised assessment that can be selected to measure function with older adults in a clinical setting.

10. Statement of research problem

At present, there is a lack of evidence as to the methodological quality and measurement properties of previous studies undertaken on the original BI. Therefore we are unable to determine whether the original BI is a valid and reliable tool to measure function with older adults.

10.1 Research aims

This research aims to undertake a systematic review of the literature on the measurement properties of the original BI. The COSMIN checklist will be used as the tool to assess the methodological quality of the included studies on measurement properties of the original BI. This research will determine which measurement properties have been investigated to date and establish the methodological quality of the existing studies. Subsequently, this research will be able to determine whether the original BI is a valid and reliable standardised assessment that measures function with older adults. The results of this study will help bridge the knowledge gap within the occupational therapy profession by providing information relating to the validity and reliability of the original BI. This will enable
occupational therapists to select the most appropriate assessment tool to measure function with older adults in their clinical setting.

10.2 Research questions

- What measurement properties of the original BI have been researched and what is the quality of these studies?
- Is the original BI a valid and reliable standardised assessment that measures function with older adults?

11. Conclusion

The proposed research will contribute to the evidence base by addressing one of the barriers occupational therapists face when using standardised assessments. Bridging the knowledge gap by providing information on the measurement properties of the original BI will better equip occupational therapists to make informed decisions related to the use of the original BI to measure function in clinical practice with older adults. This research will also enable occupational therapists to become more accountable to their profession, through the use of standardised assessment to demonstrate professional credibility when measuring outcomes and treatment effectiveness. The results from this research will be used to inform and guide clinical practice in Australia and other English speaking countries. This in turn will have the potential to decrease costs, improve patient recovery and allow for earlier discharge.
References


The measurement properties of the original Barthel Index and its applicability to measure function with older adults:

A systematic review

Author
Elizabeth Wyatt
Discipline of Occupational Therapy
Faculty of Health Sciences, University of Sydney

Contact Details
Elizabeth Wyatt
Masters of Occupational Therapy (Honours)
c/o Professor Lynette Mackenzie
Discipline of Occupational Therapy
University of Sydney
PO Box 170
Lidcombe NSW 1825
Email: ewya4745@uni.sydney.edu.au

Declarations of Interest
The author reports no declarations of interest
ABSTRACT

Aims: The original Barthel Index (BI) is a standardised functional assessment used by occupational therapists. This systematic review evaluated the measurement properties of the original BI to determine its ability to measure function with older adults.

Methods: The following databases were searched; Cumulative Index of Nursing and Allied Health (CINAHL), Embase and Medline. The methodological quality of the included studies was assessed using the COConsensus-based Standard for the selection of health Measurement INstruments (COSMIN) checklist.

Results: 3273 abstracts were screened and 5 studies included. COSMIN rated structural validity (good), cross-cultural validity (poor), hypothesis testing (fair to poor), internal consistency (poor) and responsiveness (fair). No studies investigated reliability, measurement error, criterion or content validity.

Conclusions: Multiple versions of the BI exist thus limiting the amount of research undertaken on the original BI. Further robust research is required to determine the ability of the original BI to measure function with older adults.

Key words: discharge planning, activities of daily living, measuring outcomes, standardised assessments, function, and occupational therapy.
Introduction

Throughout the world many populations are experiencing an increasing number of older adults surviving to very old age (Crews et al., 2006). In countries such as Australia, Japan, Norway, Sweden, the United States of America and the United Kingdom the number of older adults per capita has increased up to 15% of the total population (Crews et al., 2006; Cheah et al., 2011). Improvements in healthcare have contributed to older adult longevity through supporting the health related consequences of ageing (Wallace et al., 2002). Ageing can result in older adults experiencing increased frailty and incidence of falls, functional and cognitive decline and increased levels of disability or disease (Crews et al., 2006). Many of these conditions result in hospitalisation and rehabilitation, making older adults recurrent users of health care services (Giles et al., 2003). As a result, older adults are often referred to occupational therapists to improve functional capacity and support engagement in the activities and occupations necessary for daily life (American Occupational Therapy Association [AOTA], (2008).

Occupational therapists working with older adults in hospital or rehabilitation settings use standardised assessments to enhance function in self-care tasks and plan for discharge (Wales et al., 2012; Barras 2005; Parks et al., 2000; Shepperd et al., 2013). There are many different types of assessments used in clinical practice and occupational therapists are often responsible for selecting an appropriate assessment to meet their needs. Selecting an appropriate assessment can be time consuming as no assessment will measure all constructs for every client group or clinical setting (Fricke et al., 1996; Kjeken 2012). Therefore occupational therapists must consider the advantages and disadvantages of each assessment in relation to the client group and clinical setting they are working in. Then consider the
reliability and validity of the assessment in relation to their client group and clinical setting before measuring outcomes and planning therapy.

The literature suggests that some occupational therapists use standardised assessments whereas others prefer to use non-standardised assessments to measure outcomes (Unsworth 2000; Muller et al., 2011; Bowman 2006). The study by Bowman (2006) indicated the major barriers impacting on occupational therapists measuring outcomes were a lack of knowledge and skill. Other factors contributing to occupational therapists using non-standardised assessments include managerial direction, time limitations, budgetary constraints and lack of confidence (Law 1987; Bowman 2006; Atwal et al., 2007; Bowman et al., 2002; Law et al., 1989). Non-standardised assessments have not been rigorously studied or scientifically explored resulting in an inability to quantify outcomes or verify treatment effectiveness (Unsworth 2000; Law et al., 1989; Fawcett, 2007). Standardised assessments have been developed through a rigorous and scientific manner for a defined construct and population group (Fawcett, 2007). Standardised assessments allow comparison between individuals through the use of clearly defined guidelines and scoring criteria which allows objective measurement of patient change and effectiveness of treatment (Law, 1987; Fawcett, 2007; Unsworth, 2000; Law & Letts, 1989). Standardised assessments require proven reliability and validity to be used successfully with various populations, clinical settings and medical conditions (Kasner, 2006; Unsworth, 2000). Therefore, it is critical that occupational therapists understand the importance of using standardised assessments with demonstrated validity and reliability to measure outcomes within the context of their patient group and clinical setting (O’Connor et al., 2004; Fawcett 2007). The impact of incorrect measurement and interpretation of results could lead occupational therapists to provide harmful therapies,
permanent functional deficits, delay recovery and later discharge. All these factors could be detrimental to the health and well-being of individual patients.

Measuring function is an essential part of an occupational therapists role with older adults and it relies on the validity and reliability of standardised assessments. The original Barthel Index (BI) developed by Mahoney et al (1965) is a standardised assessment tool used worldwide in clinical practice today. The BI is used by occupational therapists to measure function within rehabilitation, hospital and community settings with different patient groups and medical conditions (Fricke et al., 1996; Wade 1987; Nicholl et al., 2004). The use of the original BI to measure function with older adults has not been established to date.

**The original Barthel Index**

The original BI was developed to monitor functional improvements in orthopaedic patients’ self-care and mobility tasks, before and after treatment in an in-patient rehabilitation setting (Fricke et al., 2006; Mahoney et al., 1965). The original BI measures ten activity of daily living (ADLs) that are related to functional independence or dependence in self-care tasks. These include feeding, bathing, personal toileting, and dressing, bowel/bladder control, getting on/off the toilet, locomotion, ambulation and stair climbing. The original BI uses an ordinal scale scoring system ranging from 0 to 100 points in 5-point increments. The summation of scores indicates functional capacity. A score of 100 indicates the patient is completely independent in physical function whereas a score of 0 represents a totally bedridden dependency. Assessment is performed through observation and interview delivered by allied health practitioners (Mahoney et al., 1965). Since its inception the BI has been used by occupational therapists to inform clinical decisions, plan treatment and
therapeutic interventions and measure outcomes for discharge (Mahoney et al., 1965; Atwal et al., 2008, Wales et al., 2012). The reliability and validity of the original BI for use with older adults has not been confirmed.

**Measurement properties of the original Barthel Index**

To date limited research has been undertaken on the measurement properties of the original BI with older adults (Wales et al., 2012; Law et al., 1989). So while the original BI continues to be used widely by occupational therapists, the results from assessment with older adults should be interpreted cautiously and validated with other assessments during the course of therapy until the validity and reliability of the original BI has been proven. This way occupational therapist can ensure the original BI is used appropriately in clinical practice and that the outcomes measured are truly representative of older adults’ functional capacity. Furthermore, a number of versions of the BI have been identified in the literature resulting in inconsistencies in nomenclature. Two commonly used versions developed by Collin et al (1988) and Shah et al (1989) used adapted scoring and modified guidelines. These factors impact on an occupational therapists ability to interpret the true results of previous studies when they are unable to differentiate between the original or adapted versions. Recommendations have become apparent across the literature for further research on the methodological quality of the measurement properties of the original BI (Quinn et al., 2011; Wales et al., 2012; Law et al., 1989).

This research aims to explore the measurement properties of the original BI by performing a systematic review of the literature using the COnsensus-based Standard for the selection of health Measurement INstruments criteria (COSMIN) checklist to rate the methodological
quality of the measurement properties of previous studies undertaken on the original BI. The COSMIN checklist was developed by a panel of measurement experts using an international Delphi survey for the purpose of reaching an international standard to ensure global consistency in relation to the evaluation of measurement properties in terms of study design and statistical analysis (de Vet et al., 2013). By understanding the measurement properties and limitations of the original BI occupational therapists will be able to make informed decisions when selecting an appropriate standardised assessment to measure function with older adults.

The research questions were: What measurement properties of the original BI have been researched and what is the quality of these studies? Is the original BI a valid and reliable standardised assessment that can be used to measures function with older adults?

Methods

Search Strategy

The following electronic databases were searched: Cumulative Index of Nursing and Allied Health (CINAHL), Embase and Medline were searched for relevant articles using the comprehensive search strategy. These databases were specifically targeted as they contain original research on the original BI. This research defines older adults as individuals aged 70 years or older.

Screening

Search results were exported to a reference management system where all duplicates were removed via Endnote then manual identification prior screening. All remaining abstracts
were screened against the inclusion criteria below to determine eligibility. All papers included at abstract moved to full text review. If the assessor (EW) was unsure about the eligibility of a paper, all potential papers went to the second (KW) and third (LC) reviewer for consensus on inclusion.

Papers were included in the review if: 1) the original BI was referenced and the assessment was delivered through observation and interview and scored as per the guidelines developed by Mahoney et al (1965), 2) the paper was published in English, 3) access to full text was available and 4) the paper reported on measurement properties. No age ranges were specified to ensure all articles on the measurement properties of the original BI were captured and no time limitations were set as it was decided that all literature on measurement properties was relevant. Studies were excluded if they used an alternate version of the BI which included adapted scoring, modified guidelines, alternate methods of delivery such as self-report via postal survey, telephone or proxy. As adoptions or modifications made to an assessment makes it a new assessment tool and no information related to the measurement properties of the original BI can be extrapolated to the modified versions. This research was concerned with the measurement properties of the original BI. As the original BI was written and delivered in English to other English speaking people, only papers meeting this criteria were included in this review. Studies were excluded if they did not use original data, i.e. systematic reviews.

**Evaluation of included studies**

Measurement properties have varying definitions throughout the literature and may be used interchangeably with titles such as psychometric properties leading to inconsistencies in
evaluation and reporting (Mokkink et al., 2010). The COSMIN checklist was selected for this research as it developed an international standard identifying uniformity in relation to the terms and definitions for each measurement property (Mokkink et al., 2010). By achieving agreement in the definitions and domains of measurement properties, this assures consistent terminology can be used across contexts with uniformity ensuring global consistency in relation to how measurement properties are evaluated in terms of study design and statistical analysis (Mokkink et al., 2010; de Vet et al., 2013).

The measurement properties of the included studies were evaluated and scored against the COSMIN checklist in the domains of internal consistency, reliability, measurement error, content validity, structural validity, hypothesis testing, cross cultural validity, criterion validity and responsiveness. Each domain contains 4-18 items with specific criteria to determine the methodological quality of each measurement property. The 4 point rating scale categorises each item per domain as excellent, good, fair or poor (de Vet et al., 2013). To ensure the methodological quality of the entire study, the overall rating for methodological quality equals the lowest rating received against the COSMIN checklist per item per domain (de Vet et al., 2013). If the methodological quality of the paper scores excellent or good COSMIN considers the measurement property to be reliable and the results trustworthy (de Vet et al., 2013; Mokkink et al., 2010). Similarly if the methodological quality of the paper scores fair or poor the measurement property is deemed to be doubtful and not as reliable (de Vet et al., 2013). Studies that score fair or poor have a decreased ability to derive solid conclusions about the quality of the measurement property being investigated (de Vet et al., 2013).
Generalisability and interpretability were used for data collection in order to describe the included studies characteristics and the findings. The domains of generalizability and interpretability are not scored (de Vet et al., 2013). The following table defines the measurement properties as described by the COSMIN checklist (Mokkink et al., 2010)

[Insert Table 1: COSMIN checklist definitions of measurement properties here]

COSMIN rules
The COSMIN checklist recommends that agreement is reached between the research team on the management and interpretation of subjective items. Therefore, two rules were created. The first rule recognised that a study can only be assessed once for each methodological flaw. The second rule allowed inclusion of all papers where English would be considered the primary language of the country. In this instance the assessor made the assumption that the original BI was delivered in English.

Results
After duplicates were removed 3273 abstracts were screened for inclusion. Sixty eight papers progressed to full text screening. Of these, 63 were excluded because they did not meet the inclusion criteria. The reasons are outlined in Appendix 2 and summarised in Table 3. Five articles met the inclusion criteria and were included. The search results are summarised in Figure 1.

[Insert Figure 1: Screening process here]
Table 2 provides an overview of the included studies. The mean sample size was 315, sex distribution of females to males was 59.5% and 40.5% respectively, the predominant diagnosis was stroke, and the median age of the participants was 66.5 years. 40% of the studies were undertaken in the community and 60% in hospital. The majority of studies were performed in USA, Northern Ireland and Australia. Three studies indicated by an asterisk (*) contained population groups of older adults (aged 70 years and older). These results indicate there has been surprisingly little research undertaken on the measurement properties of the original BI, which was unexpected considering the BI is used routinely within the occupational therapy profession worldwide (Fricke et al., 1996; Wade 1987; Wales et al., 2012). The results at secondary screening indicate that the many versions of the original BI are being used and referred to as the BI indicating that occupational therapists in clinical practice are not aware of the differences between the original BI and the adapted versions and how the results may impact on clinical practice.

It was interesting to note that none of the included studies were from the United Kingdom. As displayed in Table 3, of the studies excluded at secondary screening 69% were excluded due to the use of Collin and Wade scoring. On investigation as outlined in Appendix 2, 54% of the excluded studies that used the Collin and Wade scoring were from the United Kingdom. These results suggest the Collin and Wade version is preferred in the United Kingdom.
Methodological quality of the included studies

Structural validity

The de Morton et al (2008) study established structural validity performed by Rasch analysis on older adult medical patients. The Rasch model affirms an item response is the result of the interaction between a person and the items difficulty (de Morton et al., 2008). The initial results showed the original BI contained one misfitting item, namely ‘bladder’, indicating that this item is probably measuring another construct in older adults ($X^2 = 94.17, P = .00$). A secondary Rasch analysis was applied after the misfitting item was removed, however, the results still indicated that the original BI does not fit the Rasch model and is not reflective of a unidimensional scale for self-care ($X^2_{18} = 69.66, P = .00$). Additionally the findings reported that the items of feeding, stairs, using bowels and walking had disordered threshold indicating that the scoring does not accurately delineate level of difficulty. The person separation results indicated that the original BI had good internal consistency; however ceiling effects exist for a few items within in this population. Bladder and bowel items were the most misfitting items and probably relate to intrinsic health issues rather than general self-care capacity. The overall results of the de Morton study concluded that the original BI is not a unidimensional measure of self-care function for older adult medical patients and the summation of scores is not a valid indicator of function in this population, resulting in the COSMIN checklist rating the methodological quality of the de Morton study as good.
Cross cultural validity

Due to language and cultural differences that exist worldwide, investigating the impact of the measurement properties of a standardised assessment in non-English speaking countries is problematic (Beaton et al., 2000). As a result the COSMIN checklist only evaluates the quality of the translation process. The study undertaken by Loar et al (2011) was found to have poor methodological quality for a number of reasons. Only one forward and backward translation was performed using one translator per step. The COSMIN checklist recommends multiple forward and backwards translations with at least two translators per step to ensure accuracy of the translation. In this instance, Loar et al (2011) did not meet this standard. The COSMIN checklist also recommends all translators work independently of each other to detect errors or ambiguity. In the study by Loar et al (2011) it was unclear whether the translators worked independently, making the quality of the translation process questionable. The authors also failed to adequately describe the level of expertise the translators had regarding language, disease, target population and construct. Additionally, Loar et al (2011) did not describe how the differences between the original version and the translated version were resolved and did not review the translated version via a committee or pre-test in a suitable population. All these factors raise questions as to the overall quality of the translation process to accurately describe and measure function in the translated version. Therefore the results cannot be generalised. Transparency of the translation process is required otherwise any findings on the reliability and validity of the international version cannot be assumed to be an accurate reflection of the original BI.
**Hypothesis testing**

Hypothesis testing examines the construct being measured by a standardised assessment. Hypothesis testing can be used to explore the concurrent and discriminant validity between two standardised assessments that are postulated to measure the same or different constructs (Mokkink et al., 2010). For hypothesis testing authors should provide a well-constructed hypothesis to a priori with a clearly defined outcome in relation to the expected magnitude and direction of the results (Mokkink et al., 2010). Two studies by Nicholl et al (2004) and Kwon et al (2004) were included for hypothesis testing and scored poor to fair respectively for methodological quality.

The study by Nicholl et al (2004) rated poor on the COSMIN checklist as the hypothesis was unclear about whether the Community Dependency Index (CDI) was superior to the original BI. The hypothesis did not specify the expected direction or magnitude of the concurrent validity being investigated. Information on expected relationships including the magnitude should be provided so the strength of the relationships can be determined and whether or not the expected hypothesis was achieved. The results from Nicholl et al (2004) showed that the original BI and CDI are highly correlated scales \( r = 0.96 \), indicating a strong positive relationship and that they both measure the same construct within Multiple Sclerosis (MS) community setting. These results indicated that age \( r = 0.28 \) and duration of time since time of diagnosis \( r = 0.30 \) had low correlations \( r \leq 0.33 \) indicating these two factors do not influence scoring. The results showed that the original BI can discriminate between three disability groups with different levels of MS; walking unaided, walking aided and wheelchair user. The authors identified noticeable ceiling effects within the walking unaided group.
indicating that the original BI may not be sensitive enough to detect change in MS patients with higher levels of functionality.

Kwon et al (2004) investigated two hypotheses both rated fair on the COSMIN checklist. The first formulated a priori regarding the original BI’s ability to predict the Modified Rankin Scale (MRS) level within stroke rehabilitation. The second was poorly constructed and unclear with respect to the expected direction or magnitude of the relationship between the original BI and the motor component of the Functional Independence Measure (m-FIM) to predict the MRS level. This study reported on the concurrent validity of the original BI and m-FIM to measure similar constructs. Spearman correlation coefficients were reported to be $r_s = 0.9479 \ (P < 0.0001)$ indicating that the original BI and m-FIM measure similar constructs. The polytomous logistic regression method was used to determine the ability of the original BI to predict the MRS level. The results concluded that the original BI was unable to differentiate between 6 levels on the MRS but was appropriate for 5 levels. This provides evidence that the original BI is unable to sufficiently differentiate changes in higher scores and higher levels of functional ability, indicating the presence of ceiling effects within stroke rehabilitation patients.

**Internal consistency**

The study by Nicholl et al (2004) also investigated internal consistency and rated fair on the COSMIN checklist as the study failed to describe how the missing items were handled. Nicholl et al (2004) reported the internal consistency of the original BI by calculating Cronbach’s alpha coefficient $\alpha = 0.86$ which falls within the acceptable range of ($\alpha = 0.70 - 0.90$) meaning that the questions on the original BI measure constructs.
Responsiveness

Wallace et al (2002) investigated multiple ways to measure responsiveness in stroke rehabilitation 1 to 3 months post stroke and compared the responsiveness of the FIM to the original BI. Responsiveness is concerned with the ability of a standardised assessment to measure patient change over time (Mokkink et al., 2010). It is essential that authors describe the expected magnitude and direction of the change they are expecting to see in the population through the use of a clearly defined hypothesis. Otherwise readers are unable to determine if the standardised assessment has identified real change or whether change was a result of measurement error or other factors related to disease progression. While consensus has not been reached in the literature in regards to a consistent method for assessing responsiveness, the COSMIN checklist recommends that standardised assessments measuring the same construct be compared to see if they measure change by similar increments (Mokkink et al., 2010). Alternatively, authors may use more traditional statistical analysis such as effect size but should accompany such analysis with a clearly defined hypothesis (Mokkink et al., 2010).

In the study by Wallace et al (2002) Guyatt effect size and Receiver Operator Characteristic (ROC) curve analysis were performed on changing populations. Paired $t$-test, Liang standardised response mean, Kazis effect size and mixed model adjusted $t$-statistic were performed on stable populations. The mean scores for both the FIM and original BI for stable and changing populations indicated that both these standardised assessments have the ability to measure responsiveness at the same level within stroke rehabilitation. The results from ROC curve analysis slightly favoured the FIM whereas Guyatt effect size indicated that
the FIM and original BI are equivalent at measuring responsiveness in changing populations. Results from the stable population showed that all measures described above favoured the original BI as being slightly more responsive. The overall results from the study by Wallace and colleagues concluded that all forms of measurement support the ability of the original BI to measure responsiveness within the stroke rehabilitation 1 to 3 months post stroke in stable and changing populations. The study rated fair on the COSMIN checklist due to the fact that the authors did not describe why the patients deteriorated.

In summary, the COSMIN checklist identified one study of high methodological quality and four of low methodological quality. Meaning that the majority of previous research undertaken on the original BI is not methodologically sound and the results cannot be considered reliable or trustworthy (de Vet et al., 2013; Mokkink et al., 2010). The reasons for these ratings include; poorly constructed and unclear hypotheses, insufficient information about patient change and the measurement properties of the comparator instrument or lack of transparency in the translation process. Therefore the results of these studies should be viewed with caution as strong conclusions are unable to be drawn about the reliability and validity of the original BI from these results.

Of the measurement properties investigated no studies were found to include measurement error, reliability, content or criterion validity. One study by Loewen et al (1988) was considered for inclusion under the domain of reliability however was excluded due to an altered method of delivering and scoring the assessment. The original BI was designed to be delivered through observation and interview in real time in a naturalistic setting. The Loewen et al (1988) study videoed taped stroke rehabilitation patients undertaking the
assessment. These videos were then reviewed by occupational therapists and intra-rater and inter-rater reliability was assessed. As the occupational therapists scoring the assessment did not deliver the assessment this study did not meet the inclusion criteria. Additionally, the authors did not comment on whether the videos were watched in real time, paused or rewound which could result in altered scoring. Criterion validity was not expected as it requires comparison of the standardised assessment under investigation to be compared to a gold standard and currently there is no accepted gold standard to measure function.

**Application of the original Barthel Index to older adults**

As previously discussed, Table 2 described three studies, indicated by an asterisk (*), that used an older adult population with a mean age of > 70 years. Two studies investigated older adults in stroke rehabilitation and one study examined older acute medical patients. The studies by Wallace et al (2002) and Kwon et al (2004) were considered to have low methodological quality for the measurement properties of responsiveness and hypothesis testing. Both studies rated fair, therefore strong conclusions are unable to be drawn from these results. The study by de Morton et al (2008) rated good on the COSMIN checklist indicating that the methodological quality of the study was sound and the results can be deemed reliable and trustworthy. Overall due to the small about of research undertaken on the measurement properties of the original BI, at this point in time, we are unable to conclude that the original BI is a valid and reliable assessment that measures function with older adults.
Discussion

This research aimed to provide an overview of the measurement properties of the original BI and to establish if the original BI is a valid and reliable standardised assessment that measures function with older adults. Five studies were identified and the methodological quality and measurement properties of the included studies were rated using the COSMIN checklist. The results were structural validity (good), cross-cultural validity (poor), hypothesis testing (fair to poor), internal consistency (poor) and responsiveness (fair). The majority of the studies had methodological flaws in study design creating less confidence in the reliability and trustworthiness of the results. No studies investigated reliability, measurement error, criterion or content validity. As no gold standard currently exists to measure function, it is not unreasonable that criterion validity was not assessed. These results highlight the need for further robust research of higher methodological quality to be undertaken so that the reliability, validity and limitations of the original BI can be confirmed with older adults. Until further research has been done occupational therapists working with older adults should interpret the results gained from the original BI cautiously and validate the results against other assessments that measure similar constructs during the course of therapy.

Results of this research indicate that the original BI is not unidimensional and the practice of summing scores does not provide an accurate reflection of function in older adult medical patients. The results from de Morton et al (2008) indicate that some questions on the original BI measure function whereas others measure other constructs. These results are supported by Nicholl et al (2004) whose study concluded that the original BI questions contribute to constructs. Ceiling effects have been demonstrated in community based MS (Nichol et al,
2004) and stroke rehabilitation patients (Kwon et al, 2004) indicating that the original BI has limitations in discriminating patient change at high-end levels of functionality, meaning the original BI may not be sensitive enough for use in these populations. It is important for occupational therapists to understand the impact of the original BI ceiling effects when measuring function in older adults. The original BI alone is not sufficient to measure change at high-end levels of functionality. Using the original BI in conjunction with other assessments would improve the ability of occupational therapists to accurately interpret clinical change in older adults at high-end levels of functionality. The results also indicate that the original BI has demonstrated responsiveness in changing and stable stroke rehabilitation patients (Wallace et al, 2002). The poor quality of the translation process described in the study by Loar et al (2011) implies that any findings on the reliability and validity of the Mongolian version cannot be assumed to be an accurate reflection of the original BI. Therefore the ability of the Mongolian version to accurately measure function cannot be guaranteed.

Since its inception in clinical practice the BI has been reported to be the most widely used standardised assessment to measure function, however the results from this study indicate that 92% of the studies excluded at secondary screening were excluded because they used adapted or modified scoring or guidelines. Of these, 69% used Collin and Wade’s adapted scoring and guidelines (Frick et al., 1996; Wade 1987). Any change made to an assessment makes it a new tool, therefore these changes have resulted in many versions of the BI becoming available and as such these versions should be viewed separately. What makes this additionally difficult for readers of published journal articles is that all versions are known by the same name, the ‘Barthel Index’. This may have resulted in inexperienced readers
misinterpreting the results of a study, believing they were researching the original BI when in fact the study used the Collin and Wade scoring which may have produced different results to the original BI. The measurement properties of each version need to be assessed to ensure they are valid and reliable standardised assessments.

**Limitations**

The authors of the COSMIN checklist recommend when undertaking a systematic review on measurement properties two assessors independently rate the included studies prior reaching consensus on the final rating due to the subjective nature of some items (Mokkink et al., 2010). In this research one assessor (EW) independently assessed and rated the included articles against the COSMIN checklist. A second assessor (KW) was available to review any subjective items and agreement was reached prior to consensus on the final scores. The fact that one assessor rated the studies may introduce a bias, however, the assessor was trained and sought clarification from the research team as required.

Due to financial restraints there was an inability to translate papers; however this research was only interested in the original English version of the original BI. Therefore this research concentrated on papers written in English which means the results can only be applied to English speaking countries. As previously discussed unless the method of translation is of high quality and transparent, results from international versions of original BI cannot be assumed to be an accurate reflection of the original BI.
Implications for occupational therapy

This research will contribute to the evidence based practice of the occupational therapy profession by addressing one of the major barriers (lack of knowledge) to using standardised assessments for measuring outcomes in clinical practice. Bridging the knowledge gap by providing information on the importance of using standardised assessments with demonstrated validity and reliability will enable therapists to have a greater understanding of the value standardised assessments play in measuring outcomes and verifying treatment effectiveness. As a result occupational therapists will become more accountable to their profession and demonstrate professional credibility through the use of standardised assessments when measuring outcomes. This will have a beneficial impact on the healthcare system by reducing costs and allocation of resources and will provide patients with appropriate therapy that will ultimately improve recovery and decrease length of stay.

Recommendations

The studies by Nicholl et al (2004), Kwon et al (2004) and Wallace et al (2002) showed the use of poorly constructed hypotheses and had these studies used well-constructed hypotheses they may have rated higher on the COSMIN checklist. When hypotheses are well constructed readers are able to identify the intent of the research and determine if the hypothesis can be accepted or rejected. Replication of these results in higher quality studies using a more clearly defined hypothesis is also warranted. Additionally, the need for further research on the measurement properties of measurement error, reliability and content validity needs to be undertaken on the original BI to bridge the knowledge gap within the occupational therapy profession. In the meantime occupational therapists must be aware of
the limitations of the original BI and be cautious in their clinical interpretation of the results when measuring outcomes, planning therapy and discharging patients.

Due to the popularity of the Collin and Wade version of the BI in practice, this research recommends future studies be undertaken on the measurement properties of the Collin and Wade version to increase the knowledge base within the occupational therapy profession. This will enable occupational therapists to select the most appropriate assessment tool that accurately measures function with older adults.

In addition this research recommends the following name changes, the original BI by Mahoney and Barthel (1965) should be referred to as the “original Barthel Index” and the Collin and Wade (1988) version “The Barthel Index (Collin and Wade)”. These changes would ensure transparency and the ability for readers to interpret the results correctly.

**Conclusion**

This research has identified the need for further high quality robust research to be undertaken on the measurement properties of the original BI and the Collin and Wade version. Using the COSMIN checklist as a guideline for future research design, studies of higher methodological quality could be undertaken to determine whether the original BI is a valid and reliable assessment tool able to be used to measure function with older adults. Approaching research design in this manor would increase the likelihood that the methodological quality of future studies would rate higher on the COMSIN checklist and find reliable and trustworthy results on a global scale. Through the use of clearly defined hypotheses, researchers would be able to establish or refute the ability of the original BI to measure function with older adults.
When occupational therapists do not adhere to evidence based practice and continues to measure function incorrectly this has the ability to increase healthcare costs through misinterpretation of the results. This could have potentially adverse effects for some patients, for example, receiving incorrect therapy resulting in permanent dysfunction or delayed recovery and prolonged hospital stay. Additional training needs to be done to break down the barriers to using standardised assessments in clinical practice and the importance of adhering to evidence based principles.
References


Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, Bouter LM, de Vet HCW. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology 60:34-42. doi. 10.1016/j.jclinepi.2006.03.012


Table 1: COSMIN checklist definitions of measurement properties (Mokkink et al., 2010)

<table>
<thead>
<tr>
<th>Measurement Property</th>
<th>COSMIN checklist definitions of Measurement Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>Is associated with the interrelatedness of the assessments items. Internal consistency crosses into the domain of reliability in relation to measurement error</td>
</tr>
<tr>
<td>Reliability</td>
<td>Is concerned with the extent to which measurement is free from measurement error when the measure is repeated under several conditions with unchanging patients. The domain of reliability consists of overtime (test-retest reliability), by different raters on the same occasion (inter-rater reliability), or by the same rater on different occasions (intra-rater reliability)</td>
</tr>
<tr>
<td>Measurement error</td>
<td>Investigates random and systematic errors of patient scores</td>
</tr>
<tr>
<td>Content validity</td>
<td>Explores the degree to which the content of an assessment tool measures the construct it is measuring</td>
</tr>
<tr>
<td>Structural validity</td>
<td>Is concerned with the degree to which the measurements of an assessment tool accurately reflects the construct being measured</td>
</tr>
<tr>
<td>Hypothesis testing</td>
<td>Investigates the relationships between the constructs being measured and the assessment tool. Hypothesis testing can be used to explore the concurrent and discriminant validity between two assessment tools that are postulated to measure the same construct</td>
</tr>
<tr>
<td>Cross cultural validity</td>
<td>Explores the translation method of the tool and/or whether it has been culturally adapted. Both these factors influence the reproducibility of the translated version when compared to the original version of the assessment</td>
</tr>
<tr>
<td>Criterion validity</td>
<td>Compares the assessment tool to that of a nominated gold standard assessment tool</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>Investigates the ability of an assessment tool to measure change over time</td>
</tr>
</tbody>
</table>
### Table 2: Characteristics of included studies

<table>
<thead>
<tr>
<th>Author</th>
<th>Sample Size</th>
<th>Median or mean age (with standard deviation or range)</th>
<th>Distribution of sex</th>
<th>Diagnostic Group</th>
<th>Important disease characteristics (e.g. severity, status, duration) and description of treatment</th>
<th>Setting</th>
<th>Setting(s) in which the study was conducted (e.g. general population, primary care or hospital/rehabilitation care)</th>
<th>Country</th>
<th>Language in which the HR-PRO instrument was evaluated</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Loar et al., 2011)</td>
<td>168</td>
<td>Not specified (24.4% &lt; 25 yrs old 61.9% between 25-65 years of age and 13.7% &gt; 65 yrs old)</td>
<td>60.1% female 39.9% male</td>
<td>Substantial diversity in terms of age, sex, education, employment and perceived socio-economic status and observed disability.</td>
<td>Not specified. Observed disabilities were recorded however the nature and the type of disabilities were not defined.</td>
<td>City and rural community settings</td>
<td>Uvurkhangai province in Arvaikheer. Three habitats used: 1). yurts (tents) 2). city apartments and 3). rural dwellings.</td>
<td>Mongolia</td>
<td>Mongolian</td>
</tr>
<tr>
<td>(de Morton et al., 2008)</td>
<td>396</td>
<td>Mean age 78.9 years +/- 7.5 years.</td>
<td>55% female 45% male</td>
<td>Older adult medical patients aged 65 years or older</td>
<td>Older acute medical patients. Inclusion criteria - general medical patients aged 65 years or over assessed within 48hrs of admission. (Exclusion criteria - patients were admitted from a nursing home, palliative care, previously suffered a stroke, contraindicated mobilization issues or re-admitted during data collection).</td>
<td>Acute care hospital setting</td>
<td>General medical wards at an acute care public hospitals in Melbourne, Australia</td>
<td>Australia</td>
<td>English</td>
</tr>
<tr>
<td>Study</td>
<td>Sample Size</td>
<td>Mean Age</td>
<td>Gender Distribution</td>
<td>Recruitment Criteria</td>
<td>Setting</td>
<td>Data Collection</td>
<td>Setting</td>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>----------</td>
<td>---------------------</td>
<td>----------------------</td>
<td>---------</td>
<td>----------------</td>
<td>---------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>Nicholl et al., 2004</td>
<td>90</td>
<td>47.5 years +/- 11.9 years</td>
<td>76.7% female 23.3% male</td>
<td>Clinically definite, clinically probable or laboratory supported Multiple Sclerosis (MS) were recruited as part of an ongoing community based clinical study for physiotherapeutic management of lower back pain in people with MS</td>
<td>Patients home (community setting)</td>
<td>Assessments were performed in the participants home</td>
<td>Northern Island</td>
<td>Assume English because the original BI was developed in English</td>
<td></td>
</tr>
<tr>
<td>Wallace et al., 2002</td>
<td>459</td>
<td>69.7 years +/- 11.6 years</td>
<td>52.4% female 47.6% male</td>
<td>Confirmed eligible stroke as defined by the WHO criteria</td>
<td>12 hospitals in the greater Kansas City area</td>
<td>USA English</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kwon et al., 2004</td>
<td>459</td>
<td>70 years +/- 11.4 years</td>
<td>53.4% female 46.6% male</td>
<td>Confirmed eligibility stroke as defined by the WHO criteria</td>
<td>Data sourced from the Kansas City Stroke Study (Wallace et al., 2002).</td>
<td>Data analysis, review of hospital admission records, referrals from various Drs and allied health professionals and review of discharge codes</td>
<td>USA English</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Reasons for exclusion at secondary screening

<table>
<thead>
<tr>
<th>Reasons for exclusion at secondary screening</th>
<th>% papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptations or modifications to OBI (scoring or guidelines)</td>
<td></td>
</tr>
<tr>
<td>• Collin and Wade version</td>
<td>69</td>
</tr>
<tr>
<td>• Modified Barthel Index</td>
<td>2</td>
</tr>
<tr>
<td>• Others</td>
<td>2</td>
</tr>
<tr>
<td>Method of translation not described</td>
<td>15</td>
</tr>
<tr>
<td>Research was not on measurement properties</td>
<td>9</td>
</tr>
<tr>
<td>Research was not original data</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 4: Methodological quality of the included studies

<table>
<thead>
<tr>
<th>Included studies</th>
<th>COSMIN checklist domains of measurement properties</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Authors</td>
</tr>
<tr>
<td></td>
<td>Internal consistency</td>
</tr>
<tr>
<td>(Loar et al., 2011)</td>
<td></td>
</tr>
<tr>
<td>(de Morton et al., 2008)</td>
<td></td>
</tr>
<tr>
<td>(Nicholl et al., 2004)</td>
<td>POOR</td>
</tr>
<tr>
<td>(Wallace et al., 2002)</td>
<td></td>
</tr>
<tr>
<td>(Kwon et al., 2004)</td>
<td></td>
</tr>
</tbody>
</table>
**Figure 1: Screening Process**

Appendix 1: Search Strategy

Medline

Filter/ Limits: published in English

OR "consensus development conference, nih"[Publication Type] OR "practice guideline"[Publication Type]) OR ("animals"[MeSH Terms] NOT "humans"[MeSH Terms])

Embase

Manually Hit [english]/lim AND [humans]/lim AND [abstracts])

((“title of assessment tool”) AND reproducib* OR psychometr* OR clinimetr* OR clinometr* OR 'observer variation reliability' OR valid* OR coefficient OR 'internal consistency' OR (cronbach* AND (alpha OR alphas)) OR 'item correlation' OR 'item correlations' OR 'item selection' OR 'item selections' OR 'item reduction' OR 'item reductions' OR agreement OR precision OR imprecision OR 'precise values' OR 'test retest' OR (test AND retest) OR (reliab* AND (test OR retest)) OR stability OR interrater OR 'inter rater' OR intrarater OR 'intra rater' OR intertester OR 'inter tester' OR intratester OR 'intra tester' OR interobserver OR 'inter observer' OR intraobserver OR 'intra observer' OR intertechnician OR 'inter technician' OR intratechnician OR 'intra technician' OR interexaminer OR 'inter examiner' OR intraexaminer OR 'intra examiner' OR interassay OR 'inter assay' OR intraassay AND 'intra assay' OR interindividual OR 'inter individual' OR intraindividual OR 'intra individual' OR interparticipant OR 'inter participant' OR intraparticipant OR 'intra participant' OR kappa OR kappas OR 'coefficient of variation' OR repeatab* OR (replicab* OR repeated AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza* OR generalisa* OR concordance OR (intraclass AND correlation*) OR discriminative OR 'known group' OR 'factor analysis' OR 'factor analyses' OR 'factor structure' OR 'factor structures' OR dimensionality OR subscale* OR 'multitrait scaling analysis' OR 'multitrait scaling analyses' OR 'item discriminant' OR 'interscale correlation'
OR 'interscale correlations' OR (error OR errors AND (measure* OR correlat* OR evaluat* OR accuracy OR accurate OR precision OR mean)) OR 'individual variability' OR 'interval variability' OR 'rate variability' OR 'variability analysis' OR (uncertainty AND (measurement OR measuring)) OR 'standard error of measurement' OR sensitiv* OR responsive* OR (limit AND detection) OR 'minimal detectable concentration' OR interpretab* OR (small* AND (real OR detectable) AND (change OR difference)) OR 'meaningful change' OR 'minimal important change' OR 'minimal important difference' OR 'minimally important change' OR 'minimally important difference' OR 'minimal detectable change' OR 'minimal detectable difference' OR 'minimally detectable change' OR 'minimally detectable difference' OR 'minimal real change' OR 'minimal real difference' OR 'minimally real change' OR 'minimally real difference' OR 'ceiling effect' OR 'floor effect' OR 'item response model' ORirt OR rasch OR 'differential item functioning' OR dif OR 'computer adaptive testing' OR 'item bank' OR 'cross-cultural equivalence' OR 'statistical parameters'/exp OR 'psychometry'/exp OR 'questionnaire'/exp OR 'named inventories, questionnaires and rating scales'/exp OR 'outcome assessment'/exp

**Cumulative Index of Nursing and Allied Health (CINAHL)**

Limits: “abstract available”

((“title of assessment tool”) AND ((MH "Research Measurement+") OR (MH "Outcome Assessment") OR (MH "Outcomes Research") OR (reproducib* OR psychometr* OR clinimetr* OR clinometr* OR ‘item selection’ OR ‘item reduction’ OR ‘observer variation reliability’ OR valid* OR coefficient OR ‘internal consistency’ OR (cronbach* AND (alpha OR alphas))) OR ‘item correlation’ OR ‘item correlations’ OR ‘item selection’ OR ‘item selections’ OR ‘item reduction’ OR ‘item reductions’ OR agreement OR precision OR
imprecision OR ‘precise values’ OR ‘test retest’ OR (test AND retest) OR (reliab* AND (test OR retest)) OR stability OR interrater OR ‘inter rater’ OR intrarater OR ‘intra rater’ OR intertester OR ‘inter tester’ OR intratester OR ‘intra tester’ OR interobserver OR ‘inter observer’ OR intraserver OR ‘intra observer’ OR intertechnician OR ‘inter technician’ OR intratechnician OR ‘intra technician’ OR interexaminer OR ‘inter examiner’ OR intraexaminer OR ‘intra examiner’ OR interassay OR ‘inter assay’ OR intraassay))

“Assessment tool” (Replace assessment tool with one below)

- “Assessment of Motor and Process Skills” OR “AMPS” = Y
- Canadian Occupational Performance Measure OR “COPM” = Y
- “Barthel Index” = Y
- “Nottingham Extended Activities of Daily Living” OR “NEADL” OR “Nottingham extended ADL Scale” OR “Extended ADL Scale” or “EADL scale” = Y
- “Northwick Park ADL scale” OR “Northwick Park Index of Independence in ADL” OR “Northwick Park ADL Index” OR “Northwick Park Activities of daily living index” Review search terms = Y
- “Functional Status Questionnaire” OR “FSQ” = Y
- “ADL Staircase” OR “Staircase of ADL” OR “Activity of Daily Living Staircase” = Y
- “Functional Measurement Tool” OR “FMT” = Y
- “Life-Space Assessment” OR “LSA” = Y
- “Klein-Bell Activities of Daily Living Scale” OR “Klein-Bell ADL Scale” = Y
- “Disability rating index” OR “DRI”
- “Adelaide activities profile” = Y
• “Reintegration to Normal Living Index” OR “RNLI” = Y
• “Katz Index of Independence in Activities of Daily living” OR “Katz ADL scale” OR “Katz activities of daily living scale” OR “Index of Independence in Activities of Daily Living” OR “Katz index” = Y
• “Frenchay Activity Index” = Y
• Instrumental Activity Measure” = Y
• “ Groningen Activity Restriction Scale” =Y
• "Unified Parkinson's Disease Rating Scale" OR UPDRS”
• “Perceived security in daily occupations” =Y
• “Interview deterioration in daily activities in dementia” OR “IDDD” OR "interview of deterioration in daily activities in dementia"
## Appendix 2: Reasons for exclusion at secondary screening

<table>
<thead>
<tr>
<th>List of papers excluded at secondary screening</th>
<th>Country</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Batcho et al., 2012)</td>
<td>Belgium</td>
<td>Correct reference to original Barthel Index but not translated</td>
</tr>
<tr>
<td>(Cioncoloni et al., 2012)</td>
<td>107 European sited used</td>
<td>Correct reference to original Barthel Index but not translated</td>
</tr>
<tr>
<td>(Dijkstra et al., 2005)</td>
<td>Germany</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Dromerick et al., 2003)</td>
<td>USA</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Hartigan et al., 2011)</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Harwood et al., 2000)</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(van Hartingsveld et al., 2006)</td>
<td>Netherlands</td>
<td>Correct reference to original Barthel Index but not translated</td>
</tr>
<tr>
<td>(Houlden et al., 2006)</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Hsieh et al., 2007)</td>
<td>Taiwan</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Kwakkel et al., 2011)</td>
<td>Netherlands</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Loewen et al., 1990)</td>
<td>Canada</td>
<td>Outcomes study - not on MP</td>
</tr>
<tr>
<td>(McPherson et al., 1993)</td>
<td>UK</td>
<td>Comparison study - not MP of BI</td>
</tr>
<tr>
<td>(McPherson et al., 1997)</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Novak et al., 1996)</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Nyein et al., 1999)</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Post et al., 2002)</td>
<td>Netherlands</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Ranhoff et al., 1993)</td>
<td>Norway</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Richards et al., 2000)</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Sackley et al., 2005)</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Sadaria et al., 2001)</td>
<td>USA</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Sarker et al., 2012)</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Scheppers et al., 2006)</td>
<td>Netherlands</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Tennant et al., 1996)</td>
<td>UK</td>
<td>Does not meet COSMIN criteria (ordinal scale is not MP)</td>
</tr>
<tr>
<td>Authors</td>
<td>Country</td>
<td>Methodology</td>
</tr>
<tr>
<td>-------------------------</td>
<td>----------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>Hinkle et al., 2008</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>Ward et al., 1988</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>Fricke et al., 1996</td>
<td>Australia</td>
<td>Used modified versions from C&amp;W - unclear what these modifications were</td>
</tr>
<tr>
<td>Cano et al., 2006</td>
<td></td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>Celani et al., 2002</td>
<td>Italy</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>Cincure et al., 2009</td>
<td>Brazil</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>Duncan et al., 2003</td>
<td>USA and Canada</td>
<td>Comparison study - not on MP</td>
</tr>
<tr>
<td>Eichhorn-Kissel et al., 2001</td>
<td>Germany</td>
<td>Correct reference to original Barthel Index but not translated</td>
</tr>
<tr>
<td>Frihagen et al., 2008</td>
<td>Norway</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>Green et al., 2001</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>Hobart et al., 2010</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>Hsueh et al., 2002</td>
<td>Taiwan</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>Kasner et al., 2006</td>
<td>USA</td>
<td>Review of 4 scales - not original data - not MP</td>
</tr>
<tr>
<td>Kay et al., 1997</td>
<td>Hong Kong</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>Kim et al., 2003</td>
<td>Korea</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>Lindeboom et al., 2005</td>
<td>Netherlands</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>Martinsson et al., 2006</td>
<td>Sweden</td>
<td>Correct reference to original Barthel Index but not translated</td>
</tr>
<tr>
<td>O'Connor et al., 2004</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>Oveisgharan et al., 2006</td>
<td>Iran</td>
<td>No reference to original Mahoney BI (was references to already translated versions and they adapted the tool)</td>
</tr>
<tr>
<td>Richards et al., 2000</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>Rookwood et al., 1996</td>
<td>Canada</td>
<td>Comparison study - not MP of BI</td>
</tr>
<tr>
<td>Sarker et al., 2012</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>Turner-Strokes et al., 2010</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>van Balen et al., 2003</td>
<td>Netherlands</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>Reference</td>
<td>Location</td>
<td>Scoring Method</td>
</tr>
<tr>
<td>----------------------------</td>
<td>--------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>(van Bennekom et al., 1996)</td>
<td>Amsterdam</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(van Der Putten et al., 1999)</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Yang et al., 2008)</td>
<td>Hong Kong</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Sidik et al., 2010)</td>
<td>Malaysia</td>
<td>Modified the Barthel Index</td>
</tr>
<tr>
<td>(Gosman-Hedstrom et al., 2000)</td>
<td>Sweden</td>
<td>Correct reference to original Barthel Index but not translated</td>
</tr>
<tr>
<td>(Miskelly et al., 1993)</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Hsueh et al., 2001)</td>
<td>Taiwan</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Eichhorn-Kissel et al., 2011)</td>
<td>Austria</td>
<td>Correct reference to original Barthel Index but not translated</td>
</tr>
<tr>
<td>(Yasunga et al., 2007)</td>
<td>Japan</td>
<td>Correct reference to original Barthel Index but not translated</td>
</tr>
<tr>
<td>(Turner-Strokes et al., 2002)</td>
<td>UK</td>
<td>Commentary/review - not MP of BI</td>
</tr>
<tr>
<td>(Hobart at al., 2001)</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Hsieh et al., 2000)</td>
<td>Taiwan</td>
<td>Correct reference to original Barthel Index but not translated</td>
</tr>
<tr>
<td>(Jenkinson et al., 2000)</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Ali et al., 1998)</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
<tr>
<td>(Stone et al., 1994)</td>
<td>UK</td>
<td>Used C&amp;W scoring</td>
</tr>
</tbody>
</table>
Physical & Occupational Therapy in Geriatrics

Instructions to Authors

About the Journal

Aims and Scope

This comprehensive journal is recognized for its useful balance of research and clinical practice articles. For more than twenty five years Physical & Occupational Therapy in Geriatrics has functioned as a forum for allied health professionals as well as others with a focus on rehabilitation of the geriatric client to share information, clinical experience, research, and therapeutic practice. Each issue focuses on current practice and emerging issues in the care of the older client, including rehabilitation and long-term care in institutional and community settings, and innovative programming; the entire range of problems experienced by the elderly; and the current skills needed for working with older clients. Contributors consider the current methods of managing older people at home, in assisted living, alone or with families. Contributors address policy issues that affect the styles of living of older people, and discuss projects relating to research and teaching as they may affect practice in the field of gerontology.

Manuscript Submission

Manuscripts submitted to Physical & Occupational Therapy in Geriatrics (POTG) should address topics related to geriatric rehabilitation, long term care and wellness. All enquiries should be directed to the Editor.

Submissions can be made in the form of:
- Original Research;
- Case Reports;
- Systemic Reviews; and
- Theory/Perspective studies related to older adults.

Manuscripts should be submitted electronically via the journal’s online submission system, [http://mc.manuscriptcentral.com/wpog](http://mc.manuscriptcentral.com/wpog).

Please include a Microsoft Word file (.doc) cover sheet that should include all identifying information. A second Microsoft Word file should include the manuscript (abstract, text, references). Tables or figures should also be uploaded as separate documents. Authors should not include their names, telephone numbers, fax numbers or e-mail addresses inside the body of the manuscript or on any figures or tables. All identifying information will be kept confidential by the journal office.

Review Process

Manuscripts submitted to POTG undergo an anonymous review by two reviewers. Authors are emailed the reviews and a letter from the Editor summarizing the reviews and the status of the manuscript (accept, revise, rejected). Every effort is made to complete the review process in 10-15 weeks. When the recommendation is to revise, authors should resubmit the manuscript within 60 days after the revisions are requested. If the revised manuscript is not received within 60 days, the manuscript file will be closed. An extension of the deadline may be requested. Papers are frequently accepted by the Editor contingent upon changes that are mandated by anonymous specialist referees and/or members of the editorial board. If the Editor returns your manuscript for revisions, you are responsible for incorporating these revisions.
Manuscript Preparation

Spacing: Double-spaced, including endnotes and references.
Font: Times New Roman, 12 point.
Margins: Leave at least one inch margin on all four sides: set all notes as endnotes.
Page Length: Maximum 20 typed pages with the above formatting (Excluding abstract and references).
Page numbers: A header or footer on each page.

Spelling, Grammar and Punctuation: Authors are responsible for preparing manuscript copy which is clearly written in acceptable, scholarly English and which contains no errors of spelling, grammar, or punctuation.

Please be sure to be consistent in the use of abbreviations, terminology, and in citing references, from one part of your paper to another. Check the accuracy of all arithmetic calculations, statistics, numerical data, text citations and references.

Title Page
The title page should be uploaded as a separate MS Word file and include the following:
- A title that is concise and reflects the content of the manuscript
- The full name(s) of each author
- Mailing and email address of corresponding author
- Declarations and acknowledgements - please see below for further information

Manuscript: (uploaded as a separate MS Word file) should include the abstract, text and references. A header or footer with abbreviated title and page number of total (e.g., pg 2 of 7) should appear on each page.

Abstract: 100-150 words. The abstract should be structured under the following headings: Aims; Methods; Results; and Conclusions. Do not include authors’ names and affiliations on the Abstract page.

Keywords: Below the Abstract provide 3-10 keywords for index purposes.

Manuscript Style and References
Citation in the text is by author and date e.g. (McNulty 2008). Where there is more than one author include „et al.”, e.g. (McNulty et al., 2008). Parenthetical references should be placed at the end of the sentence, before the full-stop, when a resource has been used.
The list of references appears alphabetically by primary author’s last name.

Examples:
Journal article:

Chapter in a book:

Tables and Figures
Tables and figures should be uploaded electronically as separate files. Use only those illustrations that clarify and augment the text. Tables and figures must be referred to in the text and numbered in order of their appearance. Each table and figure should have a complete, descriptive title; and each table column an appropriate heading. The place at which a table or figure is to be inserted in the printed text should be indicated clearly on a manuscript: [Insert table 2 about here]

Each table and/or figure must have a title that explains its purpose without reference to the text. Please format graphs, figures etc. mindful that these will be reproduced in black & white unless the cost of colour reproduction is borne by the authors. The use of differing line types and symbols are more clearly distinguished by readers than subtle differences in colour and identical line and symbol types.

Digital files are recommended for highest quality reproduction and should be 300 dpi or higher and sized to fit on journal page. Specific permission for facial photographs of patients is required. A letter of consent must accompany the photographs of patients in which a possibility of identification exists. It is not sufficient to cover the eyes to mask identity. Please note that it is in the author's interest to provide the highest quality figure format possible. Please do not hesitate to contact our Production Department if you have any queries.

Acknowledgments and Declaration of Interest sections

Acknowledgments and Declaration of interest sections are different, and each has a specific purpose. The Acknowledgments section details special thanks, personal assistance, and dedications. Contributions from individuals who do not qualify for authorship should also be acknowledged here. Declarations of interest, however, refer to statements of financial support and/or statements of potential conflict of interest. Within this section also belongs disclosure of scientific writing assistance (use of an agency or agency/ freelance writer), grant support and numbers, and statements of employment, if applicable.

Acknowledgments section

Any acknowledgments authors wish to make should be included in a separate headed section at the end of the manuscript preceding any appendices, and before the references section. Please do not incorporate acknowledgments into notes or biographical notes.

Declaration of Interest section

All declarations of interest must be outlined under the subheading “Declaration of interest”. If authors have no declarations of interest to report, this must be explicitly stated. The suggested, but not mandatory, wording in such an instance is: The authors report no declarations of interest. When submitting a paper via ScholarOne Manuscripts, the “Declaration of interest” field is compulsory (authors must either state the disclosures or report that there are none). If this section is left empty authors will not be able to progress with the submission.

Please note: for NIH/Wellcome-funded papers, the grant number(s) must be included in the Declaration of Interest statement.

Click here to view our full Declaration of Interest Policy.