Outcome measures validated for use in stroke rehabilitation in low- and middle-income countries: A systematic review.

Nabila Schoonraad

Thesis presented in partial fulfilment of the requirements for the degree of Master of Physiotherapy in the Faculty of Medicine and Health Science at Stellenbosch University

Division of Physiotherapy

Department of Health and Rehabilitation Sciences

Faculty of Medicine and Health Sciences

Stellenbosch University

Supervisor: Prof QA Louw

Co-supervisor: Mrs G Inglis-Jassiem

March 2020

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signature: Nabila Schoonraad Date: March 2020

Abstract

Introduction:

The use of outcome measures necessary for effective clinical practice and in order to obtain reliable results in research. The commonly used outcome measures in stroke rehabilitation was developed in well-resourced high-income countries. When these outcome measures are used in a different setting, such as in low- and middle-income countries, it may require translation, cross-cultural adaptation and an evaluation of its measurement properties.

Objective:

Review the current literature reporting on outcome measures used in stroke rehabilitation that were validated for use in low- and middle-income countries.

Methods:

A comprehensive search of the following electronic databases was conducted: *Africa Journal Online; AOSIS Publishing; BioMed Central; Cochrane Library; EBSCO Africa-Wide Information & CINAHL; PEDro/Physiotherapy Choices; ProQuest; PubMed: MedLine; Sabinet African Journals; Science Direct; Scopus and Web of Science.* A unique search string was used for each database. Specific inclusion and exclusion criteria were used when considering eligibility of studies, and the reference list of included studies were searched for additional studies.

All the included studies underwent an evaluation of its quality. A self-developed data extraction sheet was used for information gathering and analysis. The studies reporting on the translation and cross-cultural adaptation process was assessed against the criteria as stated in Beaton et al., (2000). A critical appraisal tool as described by Brink and Louw (2011) was applied to all included studies in order to evaluate its methodological procedures. The reported results of statistical tests were used to interpret the psychometric properties of each outcome measure. The updated criteria for good measurement properties as reported in COSMIN (Mokkink et al., 2018) was used as a reference in this analysis.

Results:

A total of 24 studies were included in this review. Three studies took place in low income countries (Uganda and Benin). Four studies occurred in lower middle-income countries

(India, Philippines and Nigeria). The remaining 16 studies took place in upper middle-income countries (Iran, Colombia, Thailand, Brazil, South Africa, Turkey and China). After the evaluation of the methodological quality of the studies and an analysis of the psychometric properties of its outcome measures and correlation with reference standards, a final total of 23 outcome measures was recommended for use in LMICs. These outcome measures include: ABILOCO; 10MWT; BESTest; Berg Balance Scale; Postural Assessment Scale; Community Balance and Mobility scale; MiniBESTest; FIM-P; Comfortable gait speed; Maximal gait speed; Comfortable ascending stairs; Maximal ascending stairs; Comfortable descending stairs; Maximal descending stairs; Timed 'Up and Go'; Modified Ashworth Scale; Modified Modified Ashworth Scale; Persian version of the Modified Ashworth Scale; Bahasa Malaysian version of the Montreal Cognitive Assessment; Ibadan version of the Stroke Specific Pain Scale; Upright Motor Control Test (Knee Flexion subscale & Knee Extension subscale); Wisconsin Gait Scale.

Conclusion:

These outcome measures have been validated for use in lower income countries and within a specific sample population only. It is advised that clinicians and researchers consider these factors when choosing an outcome measure in the management of people with stroke. This is to ensure the measurement property of the outcome measure and thus obtain credible results.

Key words:

Stroke Rehabilitation; Outcome measure; Low- and middle-income country; psychometric properties.

Opsomming

Inleiding:

Die gebruik van uitkomsmaatreëls wat nodig is vir effektiewe kliniese praktyk en om betroubare resultate in navorsing te verkry. Die algemeen gebruikte uitkomsmaatstawwe vir beroerterehabilitasie is ontwikkel in lande met 'n hoë inkomste wat goed voorsien het. As hierdie uitkomsmaatstawwe in 'n ander omgewing gebruik word, soos in lande met 'n lae en middelinkomste, kan dit vertaling, kruiskulturele aanpassing en 'n evaluering van die meeteienskappe daarvan vereis.

Doelwitte:

Om die uitkomsmaatreëls wat tydens beroerterehabilitasie gebruik is, te hersien wat gevalideer is vir gebruik in lande met lae en middelinkomste.

Metodiek:

'n Uitgebreide ondersoek na die volgende elektroniese databasisse is uitgevoer: Africa Journal Online; AOSIS Publishing; BioMed Central; Cochrane Library; EBSCO Africa-Wide Information & CINAHL; PEDro/Physiotherapy Choices; ProQuest; PubMed: MedLine; Sabinet African Journals; Science Direct; Scopus and Web of Science. 'n Unieke soekstring is vir elke databasis gebruik. Spesifieke insluiting en uitsluitingskriteria is gebruik by die oorweging van die geskiktheid van studies, en die verwysingslys van ingesluit studies het gesoek na aanvullende studies.

Al die studies wat ingesluit is, het die kwaliteit daarvan beoordeel. 'n Selfontwikkelde data-onttrekkingsblad is gebruik vir die insameling en ontleding van inligting. Die studies wat verslag gedoen het oor die vertaal- en kruiskulturele aanpassingsproses is beoordeel aan die hand van die kriteria soos uiteengesit in Beaton et al (2000). 'n Kritiese waarderingsinstrument soos beskryf deur Brink en Louw (2011) is op alle ingesluit studies toegepas om die metodologiese prosedures daarvan te evalueer. Die gemelde resultate van statistiese toetse is gebruik om die psigometriese eienskappe van elke uitkomsmaatstaf te interpreteer. Die bygewerkte kriteria vir goeie metingseienskappe soos gemeld in COSMIN (Mokkink et al., 2018) is as verwysing in hierdie analise gebruik.

Resultate:

Altesaam 24 studies is by hierdie oorsig ingesluit. Drie studies het in lande met lae

inkomste (Uganda en Benin) plaasgevind. Vier studies het in lande met laer middelinkomste (Indië, Filippyne en Nigerië) voorgekom. Die oorblywende 16 studies het in lande met die boonste middelinkomste (Iran, Colombia, Thailand, Brasilië, Suid-Afrika, Turkye en China) plaasgevind. Na die evaluering van die metodologiese kwaliteit van die studies en 'n ontleding van die psigometriese eienskappe van die uitkomsmaatstawwe en korrelasie met verwysingstandaarde, word 'n finale totaal van 23 uitkomsmetings aanbeveel vir gebruik in LMIC's. Hierdie uitkomsmaatreëls sluit in: ABILOCO; 10MWT; BESTest; Berg balansskaal; Posturale assesseringskaal; Gemeenskapsbalans- en mobiliteitsskaal; MiniBESTest; FIM-P; Gemaklike gangspoed; Maksimum gangspoed; Gemaklike opgaande trappe; Maksimum opgaande trappe; Gemaklike trappende dalende; Maksimum dalende trappe; Tydopgestel 'Up and Go'; Gewysigde Ashworth-skaal; Gewysigde Gewysigde Ashworth-skaal; Persiese weergawe van die gewysigde Ashworth-skaal; Bahasa Maleisiese weergawe van die Montreal Cognitive Assessment; Ibadan-weergawe van die beroerte-spesifieke pynskaal; Staanmotoriese beheertoets (subskaal vir knie-flexie en onderskaal vir knie-uitbreiding); Wisconsin-gangskaal.

Gevolgtrekking:

Hierdie uitkomsmaatreëls is slegs geldig vir gebruik in lande met laer inkomste en binne 'n spesifieke steekproefpopulasie. Dit word aanbeveel dat klinici en navorsers hierdie faktore in ag neem by die keuse van 'n uitkomsmaatreël in die hantering van mense met 'n beroerte. Dit is om die meeteienskap van die uitkomsmaat te verseker en sodoende geloofwaardige resultate te verkry.

Sleutelwoorde:

Beroerterehabilitasie; Uitkomsmaatreël; Land met lae en middelinkomste; psigometriese eienskappe.

Acknowledgements

I would like to sincerely thank the following people for their contributions:

- My supervisors, Professor Quinnete Louw and Mrs Gakeemah Inglis-Jassiem, for their guidance and assistance throughout the study.
- My family and friends, for their support during this challenging period.
- My colleagues at Frere Hospital and at the Western Cape Rehabilitation Centre, for their patience with me from start to finish.

Dedications

To my patients,

your resilience inspires me.

Table of Contents

List of Figures

List of tables

List of abbreviations

BBS	Berg Balance Scale
BESTest	Balance Evaluation Systems Test
BI	Barthel Index
CAT	Critical Appraisal Tool
CCA	Cross-Cultural Adaptation
CCV	Cross-Cultural Validation
COSMIN	Consensus-based standards for the selection of health Measurement Instruments
FIM	Functional Independence Measure
ICC	Intraclass Correlation Coefficient
ICF	International Classification of Functioning, Disability and Health
LMIC	Low- and Middle-Income country
MAS	Modified Ashworth Scale
MoCA	Montreal Cognitive Assessment
MRS	Modified Rankin Scale
MMSE	Mini-Mental State Examination
NCD	Non-communicable disease
PWS	Persons with stroke
SIS	Stroke Impact Scale
USA	United States of America
WHO	World Health Organisation
WHO GHO	World Health Organisation Global Health Observatory
10MWT	10m Walk Test

Glossary

Term	Definition	Reference
Activity	Task performance.	WHO, 2002
Activity limitations	Difficulties in task performance.	WHO, 2002
Body functions	Physiological functions of body systems.	WHO, 2002
Body Structures	Physical aspects of the body.	WHO, 2002
Criterion validity	The extent of the measurement of the items on a adapted outcome measure reflects its reference test or a 'gold standard'.	Mokkink et al 2018
Construct validity	The extent of which the results of an outcome measure are consistent with hypothesis that the outcome measure accurately measures the variable to be measured.	Mokkink et al 2018
Content validity	The extent of which the items of an outcome measure reflects the variable to be measured.	Mokkink et al 2018
Cross-cultural validity	The extent of which the items on an adapted outcome measure reflects the items of the original version of the outcome measure.	Mokkink et al 2018
Environmental factors	The conditions or circumstances in which people live.	WHO, 2002
Face validity	The extent of which an outcome measure appears to reflect the variable to be measured.	Mokkink et al 2018
Impairments	Loss of a body part or function.	WHO, 2002
Index test	Outcome measure under investigation.	Brink et al 2011
Interpretability	The extent to which a subjective meaning to the results of a objective result or change in scores.	Mokkink et al 2018
Internal consistency	The extent of connection between items of an outcome measure.	Mokkink et al 2018
Measurement error	The extent of error in results that is not due to changes in the variable to be measured.	Mokkink et al 2018
Participation	Interacting in life situation.	WHO, 2002

Participation restrictions	Challenges in interacting in life situations.	WHO, 2002
Psychometric testing	Assessment of measurement properties of an outcome measure.	Brink et al 2011
Reference test	Outcome measure used as a comparison.	Brink et al 2011
Reliability	The degree to which the results are error free.	Mokkink et al 2018
Responsiveness	An outcome measure's ability to pick up change in the variable to be measured.	Mokkink et al 2018
Structural validity	The extent of which the results of an outcome measure reflects the dimensionality of the variable to be measured.	Mokkink et al 2018
Validity	The extent of an outcome measure's ability to assesses a variable.	Mokkink et al 2018

Chapter 1: Introduction

1.1 Background

The shift towards evidence-based practice transformed the physiotherapy management of stroke patients. An important aspect of implementing evidence-based practice is the assessment of objective and person-centred outcomes. The use of outcome measures are essential for effectively evaluating condition changes and ultimately the improvement of treatment in stroke rehabilitation. In addition, the use of outcome measures allows for enhanced communication between role players within the field of rehabilitation as these outcome measures are standardised.

Its reported that the use of neurological outcome measures was higher in a lower income country (India) compared with a high-income country (Canada) (Demers et al., 2019). The study also describes the common facilitators and barriers to the use of these outcome measures. Outcome measures learned in training and according to clinical guidelines were reported as facilitators for the use of outcome measures. The barriers included time constrains, own judgment for clinical reasoning when making decisions, and lack of outcome measures available (Demers et al., 2019).

Commonly used neurological outcome measures were originally developed in high-income countries and validated for use therein. Some frequently used outcome measures include the Berg Balance Scale which was developed in Montreal, Canada by Berg, Wood-Dauphinee, Williams and Maki (1992); the Functional Independence Measure (FIM) which was developed by a task force team in the USA (Uniform Data System for Medical Rehabilitation, 2012), the Barthel Index which was developed in Baltimore, Maryland, USA (Quinn, Langhorne, & Stott, 2011).

The setting in which activities are performed may influence the outcome achieved. With significantly varying settings in developed versus developing countries, using the same outcome measure in both settings may yield incorrect results. These false results may lead to an inappropriate conclusion or have an effect on clinical management with poor repercussions for patients. Therefore, in order to successfully use an outcome measure, which was developed in a specific sample population such as a well-resourced high-income country, in a different context such as a poorly resourced low-income country, certain modifications are needed.

When making these modifications to outcome measures, a systematic approach is vital for maintaining the integrity of its measurement properties and producing credible results. This involves making necessary changes to eliminate cultural differences, and an assessment of its validity and reliability before it is used in a different setting. This process is known as “cross-cultural adaptation and validation” and ensures that the outcome measure maintains its measurement properties for clinical practice and reduces the risk of introducing bias into a research study. In addition, the use of translated and cross-culturally adapted instruments allows outcomes to be compared with other areas in the world (Coster & Mancini, 2015).

The current literature available regarding outcome measures in lower income countries is sparse. A systematic review by Lima et al., (2016) showed that only eleven studies were found which evaluated the measurement properties of the post-stroke outcome measures available in Brazil. All of these outcome measures required adjustments prior to use. However, poor measurement properties and flaws in the and cross-cultural adaptation process rendered the results inconclusive. Therefore, it is essential to adhere to protocol when modifying an outcome measure for use in a different setting in order to obtain credible results.

Thus, the purpose of this systemic review is to provide an overview of the outcome measures validated for use in low- and middle-income countries (LMIC), and to describe the measurement properties of the outcome measures. The methodological quality of the studies identified in this review, will be examined. The outcome of this systematic review will provide information required to determine if the outcome measures used in LMIC are validated for use in poorly resourced populations.

1.2 Study layout

The following chapter, Chapter 2: Literature Review includes an in-depth review of the literature pertaining to various aspects of stroke and the validation of commonly used outcome measures. Chapter 3: Methodology, provides a detailed description of the methods used when performing the systematic review. Chapter 4: Results, aims to display the outcomes of the database searches, critical appraisal, and evaluation of the methodological quality of the included studies. Chapter 5: Discussion will provide an analysis of the results and compares the results with the findings from different settings and populations. Thereafter, in Chapter 6, any relevant limitations of the study will be discussed, any recommendations will be reviewed, and relevant conclusions will be drawn.

Chapter 2: Literature Review

This chapter will contain an overview of certain aspects regarding stroke and its rehabilitation. It will briefly explain the condition, its occurrence and the impact it has on persons with stroke (PWS), their caregivers, and on society. This chapter will further provide evidence for the most frequently used outcome measures used as reference standards in stroke rehabilitation within different populations and will report on its measurement properties.

2.1 Epidemiology of stroke

Non-communicable diseases (NCD), kills 41 million people each year (World Health Organization, 2019). NCDs include diabetes and hypertension which are risk factors for stroke. There are additional factors which contribute towards increased risk of sustaining a stroke. These risk factors include smoking cigarettes, sedentary lifestyles, poor dietary choices, and a family history of stroke (genetics).The effects of stroke are far-reaching and can affect people of all age-groups, gender, race and socio-economic status (World Health Organization, 2019).

Stroke is a debilitating condition resulting from a restriction of blood flow to areas of the brain. The restriction may be caused by a blockage of a blood vessel in the brain (ischemic stroke) or a rupture of one of these vessels (haemorrhagic stroke). This lack of blood flow causes tissue damage to the brain called an infarction (Han, 2018). The resulting damage to the brain gives rise to multiple life-threatening consequences.

The clinical presentation of a stroke varies according to size and location of the affected sections of the brain. Due to the specialization of each area of the brain, damage to a specific area will cause subsequent loss of the specialised neurological function. Typical symptoms include hemiparesis with sudden onset of weakness of an arm or a leg, as well as in the face and trunk, usually on the opposite side of the cerebral damage. In addition, difficulties in communicating, visual disturbances, decline in cognitive functioning and impaired balance and co-ordination are often observed after a stroke (Teasell, 2018).

A systematic review by Feigin et al., (2009) reported that the stroke incidence more than doubled in LMICs and from 2000 to 2008 the incidence has surpassed high income countries by 20%. More recently, an editorial in the Bulletin for the World Health Organization by Johnson et al., (2016) reported that 70% of strokes worldwide occur in

LMICs.

In contrast, it is reported by the World Health Organization: Health Topics (2019) that the incidence (new cases) of stroke in developed countries is declining as medical interventions advance. This reduction can be attributed to improved control of hypertension and lower smoking levels. However, the stroke prevalence (total number) continues to increase due to more people surviving a stroke (lower mortality) and the aging of the population (World Health Organization, 2019). This declining trend in the incidence of strokes is consistent with data described in the same systematic review done 10 years ago by Feigin et al., (2009) reporting a downward trend in stroke incidence within the past four decades.

2.2 Burden of stroke

Stroke is a major cause of disability throughout the world and affecting more young people in LMICs (Katan, 2018). The following sections 2.2.1-2.2.3 will briefly review the impact of stroke on the patient, the caregiver and society.

2.2.1 Patient

For those who survive a stroke, the consequences can be severe with hemiplegia, challenges with balance, impaired speech, and disturbed cognitive and visual functions (Teasell, 2018). These impairments result in a loss of simple movement, or inability to control a movement thus affecting the ability to carry out basic tasks. These activities include simple movement patterns such as rolling over in bed or transitioning from sitting to standing, as well as multi-step complex tasks such as washing and dressing. Therefore, to a certain degree, a level of care will be required after sustaining a stroke. Mobility limitations, isolation, change of role within family and society, financial strain, sense of hopelessness and despondency are described as some of the experiences of people living with stroke (Maleka et al., 2012).

2.2.2 Caregiver

The consequences of stroke limit the PWS's ability to perform tasks as they were done before. In most cases, the PWS will require assistance from a caregiver to perform these tasks. Caregiving is defined as assistance provided by family and friends in order to carry out tasks (Pont et al., 2018). The level of care can range from supervision only to constant care throughout the day and at night. Therefore, there is a corresponding burden placed on

those who provide care. Caregiving negatively influence the physical, financial and emotional well-being of informal caregivers (Gbiri, Olawale & Isaac, 2014).

2.2.3 Society

The World Health Organization: Health Topics (2019) reports that annually 15 million people suffer a stroke, of which five million are left permanently disabled. These PWS place a strain not only on their family but also within society. It is also reported that stroke is occurring in a younger population comprising of 10-15% of all stroke patients (Smajlovic, 2015). There is a substantial economic impact on PWS when they are rendered disabled prematurely during their productive years. This impact also limits the productive force of an area and negatively affects the economic growth of a country. The burden of stroke results directly from lower levels of productivity and increased expenditure on healthcare, on both a personal and governmental level (Smajlovic, 2015).

2.3 Rehabilitation after stroke

With this tremendous strain placed on the PWS and their caregivers, the focus of rehabilitation is directed at decreasing the burden of care, regaining independence and achieving goals directed at reintegration into home, community and/or the workplace. There is a wide variety of interventions and techniques which can be used in the rehabilitation of a person with stroke. The treatment program is best individualised depending on patients' goals, and considering the severity of the condition, and each person's ability to recover (Ntsiea, 2019). Therefore, stroke rehabilitation requires specialized care from trained professionals and is best used in a multi-disciplinary team approach which includes a doctor, nursing staff, physiotherapists, occupational therapists, speech therapists, social workers, psychologists and dieticians.

A study by Cunningham and Rhoda (2014) demonstrates the positive influence of physiotherapy in the rehabilitation of stroke patients. A number of the participants found therapy to be a facilitator in their participation and stated that "The therapy helped me to walk, I can go to church again, now that I can walk alone." Similar findings were presented in a Polish study by Michalczak et al., (2017).

2.4 Stroke rehabilitation in low- and middle-income countries

It's shown that exercise and cognitive training interventions improved functional outcomes post stroke in LMICs (Dee, Lennon and O'Sullivan, 2018). This process of rehabilitation

may be prolonged and be very costly. In LMICs, the lack of resources negatively impacts the recovery of stroke patients. These constraints are further reduced by limited access to healthcare, poor infrastructure in the more rural areas, and lack of staffing. Rhoda et al., (2014) concluded improved outcomes in a well-resourced German rehabilitation centre for gross motor and upper limb function after stroke when compared to poorly-resourced outpatient services in South Africa. The German participants had better sitting balance, perform transitions, carry out transfers and walk independently with faster recovery of upper limb function.

2.5 Value of outcome measures in rehabilitation

Evidence-based practice is described by Veras et al., (2016) as a field of study, research, and practice where evidence guides clinical decisions in which ethical principles forms a base for professional practice. Without evidence-based practice guiding the care of patients, severe consequences can arise. These consequences includes implementing treatments which may be ineffective or even harmful. One of the cornerstones to implementing evidence-based practice is to maintain objectivity with the use of outcome measures. An outcome measure allows an unbiased evaluation of condition status, to note improvement/recovery, and to evaluate the response to rehabilitation intervention. Therefore, the use of outcome measures is crucial to maintain objectivity. Thus, a key aspect for effective rehabilitation and research is the use of outcome measures.

A study by Inglis, Faure and Frieg (2008) measured the use of outcome measures by physiotherapists belonging to the South African Society of Physiotherapy. Of the 168 participants, 84% used outcome measures regularly, but predominantly, impairment-related measures were used. In comparison, a study found that only 48% of a sample of physiotherapists (n=456) belonging to the American Physical Therapy Association used standardised outcome measures (Jette et al., 2009). The previously mentioned study by Inglis, Faure and Frieg (2008) found the efficiency in clinical practice (82%) and evidence-based practice (15%) were major contributing factors in the use of outcome measures. This study further showed that lack of time and sufficient knowledge in the use of outcome measures were obstacles to its use. Similar themes were raised in the study by Jette et al., (2009) stating that a great proportion of users believed that using outcome measures enhanced communication and directed care of their patients. Furthermore, reasons for not using outcome measures included length of time to complete and analyse the data, and difficulty for patients in completing them independently.

Outcome measures can also be categorised in the framework of the WHO International Classification of Functioning, Disability and Health (ICF). The ICF provides a multi-dimensional framework for health and disability suited to the classification of outcome instruments. (WHO, 2002). The ICF is made up of three domains and two contextual factors. The domains are (i) Body Structure and Function (ii) Activity Limitations, and (iii) Participation restrictions. The contextual factors include (i) environmental factors, which need to be considered in the rehabilitation of patients, and (ii) personal factors which affect the rehabilitation process. These contextual factors can be made up of various facilitators or barriers such as home environment, co-morbid diseases, family support network, etc.

2.6 The psychometric properties of an outcome measure

The Consensus-based standards for the selection of health Measurement Instruments (COSMIN) is an international initiative made up of a team of researchers with expertise in epidemiology, psychometrics and qualitative research. It assists with the selection of outcome measures in both a research and clinical setting. The COSMIN taxonomy, as described by Mokkink et al., (2018) presents the different domains of the psychometric properties of an outcome measure. These domains have different divisions. The domain of reliability has the following divisions: inter-rater reliability, intra-rater reliability, test-retest reliability, internal consistency and measurement error. The domain of validity has the following divisions: criterion validity, content validity, which is made up of face validity, and construct validity which is made up of structural validity, hypothesis testing, and cross-cultural validity. The remaining two domains are responsiveness and interpretability.

The COSMIN manual by Mokkink et al., (2018) defines these concepts as follows:

Validity domain:

- Criterion validity: The extent of the measurement of the items on a adapted outcome measure reflects its reference test or a ‘gold standard’.
- Construct validity: The extent of which the results of an outcome measure are consistent with hypothesis that the outcome measure accurately measures the variable to be measured.
- Content validity: The extent of which the items of an outcome measure reflects the variable to be measured.
- Cross-cultural validity: The extent of which the items on an adapted outcome measure reflects the items of the original version of the outcome measure.

- Face validity: The extent of which an outcome measure appears to reflect the variable to be measured.
  Structural validity: The extent of which the results of an outcome measure reflects the dimensionality of the variable to be measured.
- Validity: The extent of which an outcome measure assesses the variable it intends to measure.

Reliability domain:

- Internal consistency: The extent of connection between items of an outcome measure.
- Measurement error: The extent of error in results that is not due to changes in the variable to be measured.

Other domains:

- Interpretability: The extent to which a subjective meaning to the results of an objective result or change in scores.
- Responsiveness: An outcome measure's ability to pick up change over time in the variable to be measured.

The values reported on for each of these criteria has comparative values against which it can be interpreted. See Chapter 3 section 3.12.3 for the evaluation criteria for these measurement properties.

2.7 Overview of commonly used 'gold standard' outcome measures in stroke rehabilitation

Some widely used outcome measures can be considered to be a reference standard within the research context. These are also referred to as the 'gold standard' and are frequently used in stroke rehabilitation. These include the Modified Ashworth Scale (MAS), Mini-Mental State Examination (MMSE), Barthel Index (BI), Berg Balance Scale (BBS), Functional Independence Measure (FIM), Ten Metre Walk Test (10MWT), Stroke Impact Scale (SIS) and the London Handicap Scale (LHS). The remaining portion of this literature review will look at these outcome measures and examine research done regarding its psychometric properties.

2.7.1 Body Function and Structure related reference standard outcome measures

Modified Ashworth Scale(MAS)

The MAS is an adapted version from the original Ashworth Scale which was used in Multiple Sclerosis. Bohannon and Smith (1987) undertook this investigation in the USA and described the MAS as a five-point measure for the resistance during passive muscle stretching. The MAS has sufficient content validity as demonstrated by Min et al., (2012) and sufficient convergent validity with the Fugl-Meyer, Electromyography, Box-Block Test, Active range of movement, grip strength and the Pendulum test (Katz, Rovai, Brait & Rymer, 1992; Lin & Sabbahi 1999). Inter-rater reliability of the MAS was evaluated by Gregson et al., (2000) and produced adequate to excellent results. The MAS was also tested on elbow flexor muscle spasticity and the investigators agreed on 86.7% of their scoring. (Ansari et al., 2009). The MAS is a measure that is quick to administer and no training is necessary and was already translated into seven languages including Chinese (simplified), French, German, Italian, Japanese, Korean and Spanish.

Mini-Mental State Exam (MMSE)

The MMSE is a brief outcome measure used as a screening tool for the presence of cognitive impairment that was developed in the USA to assess for dementia in psychiatric setting (Folstein, Folstein, & McHugh, 1975). The MMSE has 11 items and when investigated for construct validity and has shown sufficient correlations with the Barthel Index as well as two depression outcome measures in an acute stroke population (Agrell & Dehlin, 2000). In reliability testing, the MMSE showed sufficient internal consistency (McDowell, Kristjansson, Hill & Hebert, 1997) and a inter-rater reliability kappa of 0.63 (Dick et al., 1984). The original version is freely available and no training is required to administer but keeping in mind that age, education and socioeconomic background can introduce bias in MMSE results (Mungas et al., 1996). There are over 50 authorised translations of the MMSE available.

2.7.2 Activity Limitation related reference standard outcome measures

Barthel Index

The BI was developed by the MAPI Research Trust in Lyon, France. Hsueh, Lin, Jeng and Hsieh (2002) showed that the Barthel Index has sufficient correlations with the FIM motor subscale, has sufficient levels of agreement between raters and internal consistency. The

outcome measure is freely available and requires no prior training. The BI was translated into Chinese, Danish, French, German and Korean.

Berg Balance Scale (BBS)

The Berg Balance Scale consists of 14 activities that evaluates balance and was developed in Canada. Mao, Hsueh, Tang, Sheu and Hsieh, 2002 demonstrated that the Berg Balance Scale has sufficient concurrent validity with the Fugl-Meyer; the Postural Assessment Scale for Stroke Patients and the Barthel Index. The BBS has sufficient test-retest reliability as tested by Liston and Brouwer, (1996); Hiengkaew, Jitaree and Chaiyawat (2012) and Flansbjer et al., (2012). Sufficient intra- and inter-rater reliability was also demonstrated by Mao, Hsueh, Tang, Sheu and Hsieh, (2002); and Berg, Wood-Dauphinee and Wiliiams (1995). The BBS does not require prior training to administer and can be completed within 15 minutes. The BBS was tested in various conditions including stroke, and translations were made into seven different languages, none of which are native to lower income countries.

Functional Independence Measure (FIM)

The FIM is an outcome measure which indicates a level of disability on a 7-point ordinal scale. It consists of evaluation of the performance of various functional tasks. The FIM has undergone extensive psychometric testing and the Canadian Partnership for Stroke Recovery (2019) reports numerous studies indicating sufficient reliability, and further states that it shows excellent correlations with the Barthel Index, Modified Rankin Scale and Disability Rating Scale. However, the FIM has a financial implication as it is required to be administered by a licensed evaluator and scored by consensus within the MDT.

Ten meter walking test (10MWT)

The 10MWT is a performance measure used to calculated walking speed. The initial documented use of a ten-meter walking test was used to evaluate recovery of walking ability after a stroke (Wade et al., 1987). The 10MWT showed sufficient construct validity with the Timed ‘Up and Go’ (Flansbjer et al., 2005) and the Berg Balance Scale (Wolf et al., 1999). It also has sufficient test-retest, inter-rater, and intra-rater reliability (Collen, Wolf & Bradshaw, 1990). The 10MWT is a freely available gait assessment and only a stopwatch and a clear pathway is required to administer the test.

2.7.3 Participation Restriction related reference standard outcome measures

Stroke Impact Scale (SIS)

The SIS is made up of eight domains and has 59 items and is used for an evaluation of health status after a stroke and was developed in the USA (Duncan et al., 1999). The SIS was shown to have sufficient construct validity and reliability, and can differentiate severity of stroke (Duncan et al., 1999). The SIS has two modified variations and is available in over ten different languages.

London Handicap Scale (LHS)

The LHS is a patient reported outcome measure used to assess the effect on functional ability. It is based on the ICF framework developed by WHO. The LHS was confirmed to be valid and reliable by Harwoord, Rogers, Dickson and Ebrahim, (1994). The outcome measure consists of 6 items scored along a 6-point interval scale. The LHS is free to use, quick to administer and no prior training is required.

2.8 Adapting outcome measures for use in a different setting

When researchers or clinicians are required to modify an outcome measure to fit a new context, it is required to undergo an adaptation process. Published guidelines such as described by Beaton et al., (2000) can be used as reference to maintain equivalence between the index and reference outcome measure. The outcome measure is required to undergo two processes (i) translation and cross-cultural adaptation, and (ii) validation. Thereafter, a test of measurement properties is done to ensure that the target version reflects the original version to ensure the integrity of the outcome measure.

2.8 Conclusion

This section provided a brief overview of the literature available regarding frequently used outcome measures in stroke rehabilitation. It demonstrates the psychometric properties of a few of the ‘gold standard’ outcome measures as tested in stroke samples. These reference standard outcome measures have undergone extensive evaluation and are shown to be valid and reliable. It is important to note that these outcome measures were developed in high income countries and their initial target population was not always people with stroke. The following chapter will explore the processes which will be used when evaluating the studies included in this study.

Chapter 3: Methodology

The procedures followed in this systematic review will be explored within this section. This chapter will also present the research design, research question, project aims, research objectives, inclusion and exclusion criteria, as well as ethical considerations of the project.

3.1 Research design

A systematic review procedure was followed for this study design and a descriptive analysis of the results obtained from the included studies.

3.2 Research question

Which outcome measures for stroke rehabilitation have been validated for use in low- and middle-income countries?

3.3 Project aim

The purpose of this study was to systematically search the current literature reporting on the validation of outcome measures used in stroke rehabilitation within low- and middle-income countries. (See Appendix A: The World Bank Classification for list of countries investigated.)

3.4 Research objectives

The primary objectives of this review were to:

- Systematically search the current literature reporting on the validation (including but not limited to face validity, content validity, construct, structural, longitudinal validity) of outcome measures used in stroke rehabilitation within low- and middle-income countries.
- Describe the type of outcome measures which have been validated for use in the rehabilitation of adult stroke patients in low- and middle-income countries.
- Describe the populations and countries (geographic and socio-demographic characteristics) within which outcome measures were validated in stroke rehabilitation.
- Report on the type of validation and current evidence for validation of outcome measures used in stroke rehabilitation within low- and middle-income countries.

Secondary objectives of this review were to:

- Critically appraise the methodological quality of these studies.
- Describe the psychometric qualities of these outcome measures.
- Where applicable, describe the cross-cultural adaptation of outcome measures.

3.5 Ethical clearance

The authors declare that they have no affiliations with or financial involvement in any organisation with a direct financial interest in the matter or resources used in this study. An electronic application was made on 28 August 2019 to the Health Research Ethics Committee of the University of Stellenbosch. A summarised version of the protocol for this systematic review was submitted along with other required documents. The application was approved via expedited review procedures on 11/09/2019. The study was issued a project ID 10703 and an Ethics Reference Number X19/08/032. See Appendix B.

3.6 Criteria for considering studies

Inclusion Criteria

Types of studies:

- Studies reporting on the validity and reliability testing of outcome measures.
- Studies published from and including 1990, to present.

Types of participants:

- Participants with a diagnosis of cerebrovascular accident (stroke).
- Only studies on human subjects were included in this systematic review.
- Studies that included both male and female participants were considered.
- Studies reporting on participants older than the age of 18 years were considered.

Types of outcome measures:

- Impairment, activity and participation measures as administered by a therapist
- Impairment, activity and participation measures as reported by a participant.

Exclusion Criteria

- Participants in a study who received any intervention other than physiotherapy management, such as surgical intervention.
- Participants in a study presenting with any neurological condition other than a cerebrovascular accident (stroke).
- Studies reporting on validity and reliability testing of an outcome measure in a

variety of conditions in addition to stroke.
- Studies reporting on participants that are still considered medically unstable and not fit for rehabilitation were excluded.

3.7 Search terms

Searches within the databases listed below in section 3.8 were performed with the selected keywords; custom-designed search strings (combination of keywords), and filters in order to eliminate potentially irrelevant information. For each individual database, a specific search strategy was developed using various database operators such as Boolean terms, phrase searching, wild cards and subject headings.

3.8 Search strategies

The first reviewer (N.S.) was responsible for conducting the searches and selecting eligible studies. The following databases were searched: *Africa Journal Online (AJOL); AOSIS Publishing; BioMed Central; Cochrane Library; EBSCO Africa-Wide Information & CINAHL; PEDro/Physiotherapy Choices; ProQuest; PubMed: MedLine; Sabinet African Journals; Science Direct; Scopus and Web of Science.*

The first researcher (N.S.) thoroughly documented all results and processes. All potential articles were screened for selection according to the inclusion and exclusion criteria as listed in section 3.6. Keywords included in the searches were: *stroke; cerebrovascular accident; assessment; outcome measure; outcome assessment; test; physiotherapy; physical therapy; rehabilitation; stroke rehabilitation; neurological rehabilitation; validity; reliability; psychometric*. The search strategies that have been specifically designed for each database are described in Appendix C.

3.9 Selection of studies

Titles of the studies identified in the database searches were independently scanned for eligibility by the first reviewer (N.S.). The abstracts of these studies were screened against the inclusion and exclusion criteria, and any duplicates were then removed. Where multiple studies were found reporting the same data, only the earliest articles were included. Thereafter, the full texts were further assessed of eligibility. Disagreements were resolved by discussion amongst the first and second (G.I.J.) and/or third reviewers (Q.L.) and inclusion of studies into this review was decided by consensus. Secondary searching (namely PEARLing) of the reference lists of included studies was conducted in order to

identify additional relevant studies.

3.10 Data extraction method

A custom-developed data extraction form, created in Microsoft Excel version 14.6.6 (160626), was used to capture extracted information from included studies. The following basic descriptive data was extracted from the included studies: author(s), study title, publication year, country of publication, study aim, study type, outcome measure(s) studied, study setting and its World Bank classification, participation description, sample age, sample gender and sample size. The type of validation conducted, and/or validation processes (for example as per the Rasch model) employed in selected studies, were also identified. The psychometric properties of the OM were analysed based on the COSMIN taxonomy of measurement properties (Mokkink et al., 2018). Data regarding the cross-cultural adaptation process was also extracted based on the criteria developed by Beaton et al., 2000. See Appendix E for the data extraction form. Data extracted were cross-checked for completeness and accuracy.

3.11 Data analysis

A descriptive analysis was performed on the data obtained from the studies included in the review. The extracted data were summarised narratively using text and tables. The outcome measures were briefly described in respect of their measurement properties and appropriateness for use in poorly resourced settings. In addition, the methodological quality of these studies was also discussed.

3.12 Study appraisal

3.12.1 Appraisal of methodological procedures

The methodological quality of included studies was reviewed by the first reviewer using criteria as described by Brink and Louw (2011). See Appendix D for the Brink and Louw (2011) critical appraisal tool. This critical appraisal tool is made up of 13 items, of which five items are directed at both validity and reliability studies, four items are directed at validity only studies and four items are directed to reliability only studies. The scoring for these 13 items can be ‘yes’, ‘no’ or ‘not applicable’. The 13 items assesses the following:

- Item 1: Was the participant sample described in detail?
- Item 2: Was the competence of index test rater described?

- Item 3: Was an explanation of the reference standard provided?
- Item 4: Was inter-rater blinding ensured?
- Item 5: Was intra-rater blinding ensured?
- Item 6: Was randomisation of the test order ensured?
- Item 7: Was the interval between the index and reference standard tests sufficient?
- Item 8: Was the stability of the variable ensured between test periods?
- Item 9: Was the reference standard included in the index test?
- Item 10: Was the index test procedure described in detail?
- Item 11: Was the reference standard test procedure described in detail?
- Item 12: Was any participants unaccounted for (withdrawals)?
- Item 13: Was the psychometric test applicable?

3.12.2 Appraisal of translation and cross-cultural adaptation procedures

In addition, studies that reported on the translation and cross-cultural adaptations of outcome measures were also appraised according to the Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures by Beaton et al., 2000. This published guideline to cross-cultural adaptation consist of the following six stages:

Stage 1: Translation

The initial translation stage requires two forward translations of the outcome measure be prepared. This is to identify any discrepancies in wording. It is advised that both the translators' mother tongue is the target language, and that one of the translators should have background into the context of the outcome measure. The other translator should be naïve to the topic and thus will be less influenced by academics.

Stage 2: Synthesis

This stage assesses the development of the common translated version of the outcome measure. In this stage it is important that discrepancies are resolved by consensus in order to maintain the integrity of the wording used in the outcome measure.

Stage 3: Back Translation

Stage three considers the back translation of the outcome measure into its original language. This is done in order to ensure that the two versions reflect the same content.

Stage 4: Expert committee review

Beaton et al., (2000) describes stage four as the expert committee review. It is advised that

the committee consists of methodologists, health professionals, language professionals, and the back-and-forward translators. The focus of the committee is to synthesize the pre-final version of the outcome measure which will be used in the next stage.

Stage 5: Pretesting
During stage five, testing of this pre-final version occurs and is ideally tested on a population sample size between 30-40. The aim is to evaluate the content validity of the outcome measure.

Stage 6: Submission and Appraisal of all written reports by developers/committee.
The final stage, stage six, occurs throughout the adaptation process and allows for all necessary correspondence with the developers of the original version, and appraisal throughout the process.

3.12.3 Appraisal of statistical methods used

The results of the validity and reliability assessment done in the included studies were compared against three various reported criteria. In the COSMIN manual by Mokkink et al., (2018) pages 28-29, which includes a table labelled ‘Updated criteria for good measurement properties’ and which consists of reference criteria for test-retest reliability, intra- and inter-rater reliability, internal consistency, measurement error, criterion validity and structural validity. See Table 3.1. below for the complete list of each evaluation criteria for the various measurement properties that were assessed across the included studies.

Table 3.1- Evaluation criteria applied to the results of psychometric testing

Measurement property	Type	Commonly used statistical methods	Criteria	
Reliability	Test retest; Inter-rater; Intrarater	Intra-class correlation coefficient (ICC); weighted kappa [k(w)]; unweighted kappa statistics (k); Pearson's coefficient, Spearman's coefficient (r); Kendall's coefficient of concordance (W); Kendall tau-b; Cohen's Kappa	≥0.70	Sufficient
			Not reported	Indeterminate
			<0.70	insufficient
	Internal consistency	Cronbach's alpha statistics (a) Scale reliability	≥0.70	Sufficient
			Criteria not met	Indeterminate
			<0.70	insufficient
	Measurement error	Standard error of measurement	SDC or LoA < MIC	Sufficient
			MIC not defined	Indeterminate
			SDC or LoA > MIC	Insufficient
Validity	Criterion validity	Correlation with 'gold standard' Spearman rank order correlation (r) Area under curve (AUC)	≥0.70	Sufficient
			Not reported	Indeterminate
			<0.70	insufficient
	Structural validity	Rasch analysis	Comparable measure >0.95	Uni-dimensionality
			Residual correlation <0.20	Local independence
			Scalability >0.30	Monotonicity
			In/out-fit MnSq: ≥0.5 & ≥1.5; z: > -2 & <2	Model fit

Legend: SDC = smallest detectable change; LoA = Limits of agreement; MIC = minimal important change; MnSq = Mean Square;

3.13 Conclusion

The Brink and Louw (2011) critical appraisal tool, Beaton et al., (2000) guidelines for cross-cultural adaptation, and the COSMIN (Mokkink et al., 2018) evaluation criteria for statistical results will be used to assess the included studies. Any deviations from these assessment criteria in the study which can influence the credibility of the results will be taken into consideration.

Chapter 4: Results

The results of the database searches will be provided in this section. An outline of the search process will be presented, as well as a brief description of the included studies. The methodological quality of these studies will also be explored. This is important to note when considering the significance of the results and to develop accurate conclusions. The methodological quality assessment of the studies will consider the procedures followed during cross-cultural adaptation and/or validation of the measurement properties of the outcome measures under investigation.

4.1 Search procedure and results

Two reviewers (N.S. and G.I.J.) screened the lists of electronic databases accessible on the University of Stellenbosch Library website. (http://library.sun.ac.za/en-za/Pages/Home.aspx). The initial searches resulted in a collective total of 804 hits from all the databases. See Figure 4.1. below for the total number of hits per database. After screening titles and/or abstracts, 698 studies were excluded based on the inclusion and exclusion criteria. This resulted in 106 studies which were considered for eligibility. The full text of these studies were further assessed against the inclusion and exclusion criteria. An additional 64 studies were excluded for the following reasons:

- Validation of an outcome measure was not the aim of the study: five studies
- Different study population: 55 studies
- Medically unstable sample: four studies

After scanning the remaining 42 studies, 20 duplicate studies were excluded. An additional two studies were included after searching the reference lists of the remaining 22 studies. When the selection process was completed, a final total of 24 studies were included in this systematic review. See Figure 4.1. below for a flow diagram to illustrate the selection process.

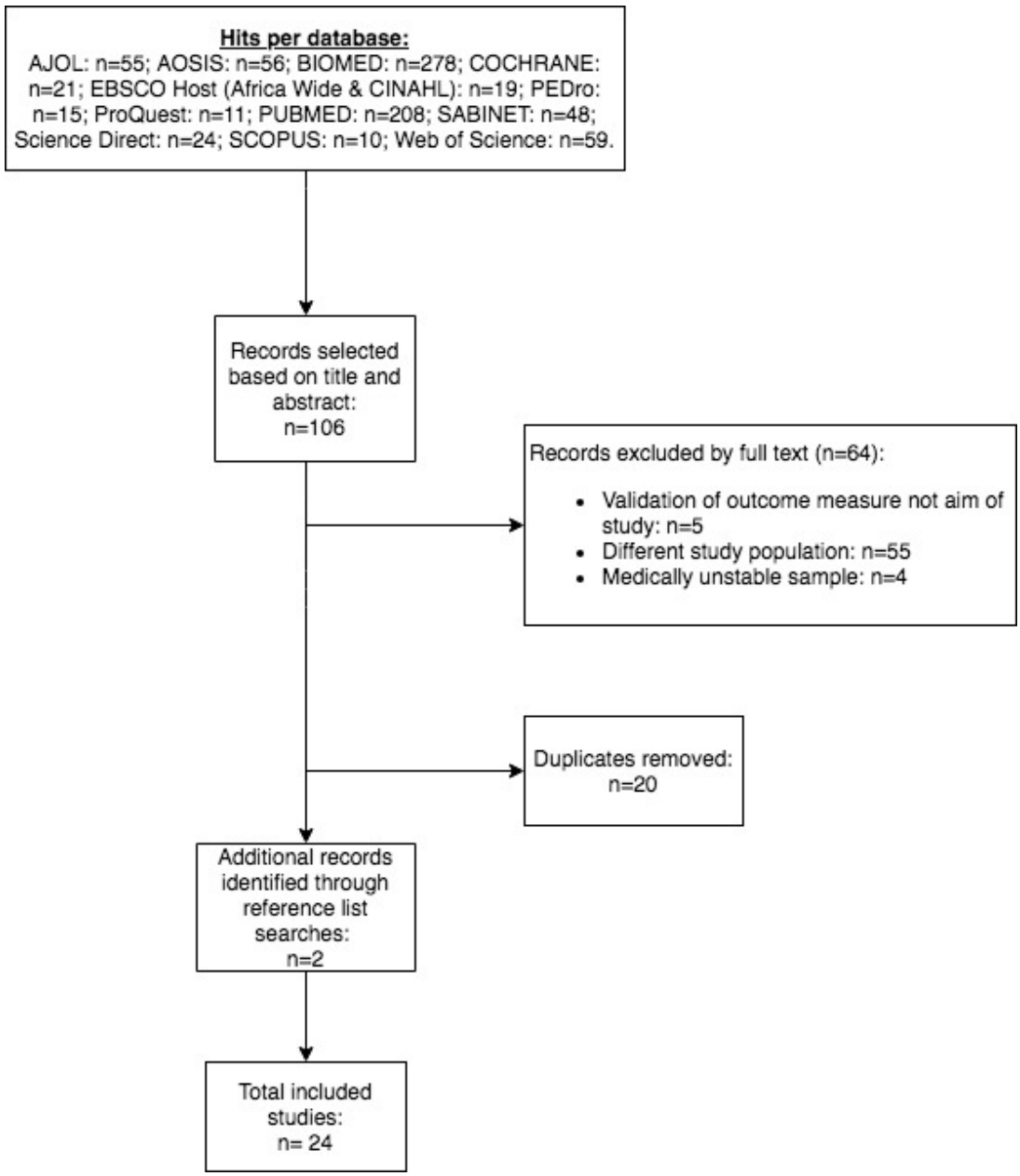

Figure 4.1 - Flow diagram of selection process of included studies

4.2 Geographical description of the included studies

Seventeen of the 24 studies included in this systematic review took place in the following upper middle-income countries: Brazil, China, Colombia, Iran, South Africa, Thailand, Turkey, and Malaysia. Four studies took place in lower-middle income countries including India, Nigeria and Philippines. Three studies took place in low-income countries namely Benin and Uganda. See Figure 2 below for a graphical representation of the various locations reported in included studies. This also serves as colour-coded legend for all tables in this chapter indicating the country's economic status according to the World Bank Classification (2018).

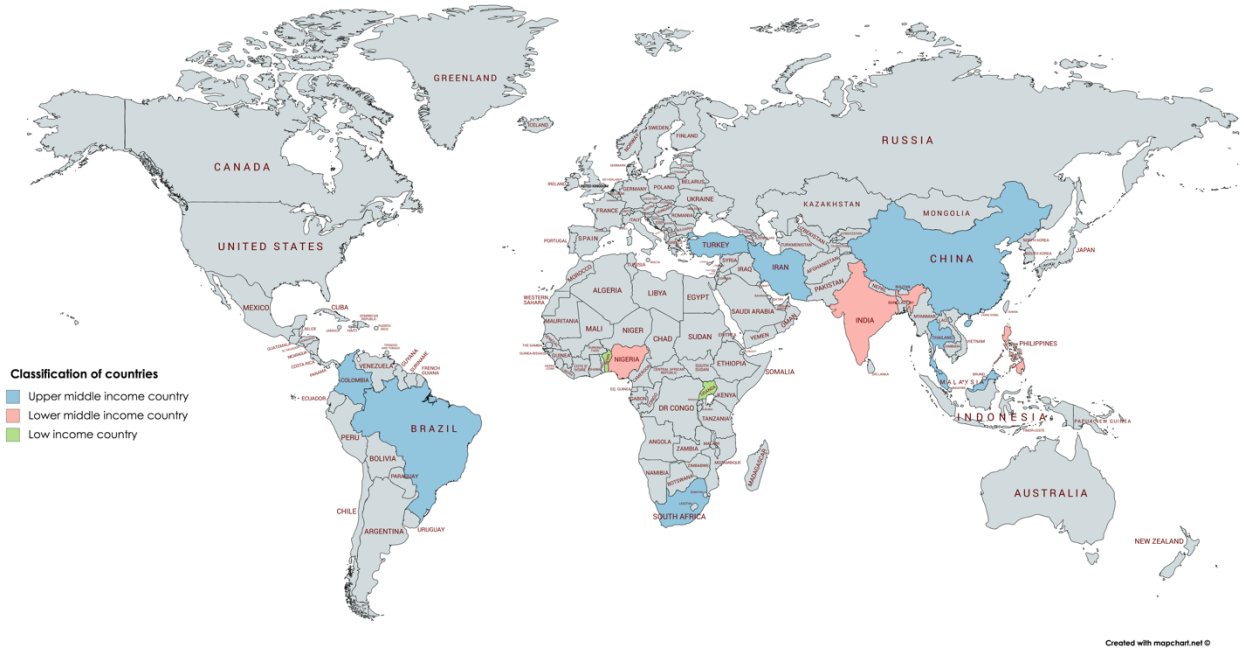

Figure 4.2 - Location of study populations

4.3 Overview of studies included in the review

Tables 4.1 to 4.3 below further describe the included studies in terms of study aims, sample characteristics, study setting, outcome assessed and which ICF domain the outcome measure represented.

Overall, only one study (Fallahpour et al., 2011) included an outcome measure exclusive to the participation domain of the ICF. Three studies (Hale et al., 1998; Kamwesiga et al., 2016; Diwan et al., 2018) covered all domains of the ICF, but the bulk of the outcome measures focused either on body structure and function or activity limitation domains. Nine studies (Ansari et al., 2008; Ansari et al., 2009; Kaya et al., 2011; Ansari et al., 2012; Li et al., 2014; Sahathevan et al., 2014; Osundiya et al., 2016; Ostrofsky et al., 2016; Barbosa et al., 2019) assessed body structure and function and 11 studies (Hamzat et al., 2009; Faria et al., 2011; Kurtais et al., 2011; Hiengkaew et al., 2012; Chinsongkram et al., 2014; Yaliman et al., 2014; Naghdi et al., 2016; Oveigharan et al., 2016; Kamwesiga et al., 2016; Gelisanga et al., 2019; Niama Nata et al., 2019) assessed activity limitations respectively.

The contextual factors of environment was addressed in some studies, but none assessed personal factors affecting rehabilitation.

Table 4.1, which has been colour-coded as blue to represent upper-middle income countries, describes the17 studies published in the span of 20 years ranging from 1998 to 2018. Only three of these studies were published in the first 10 years (1998-2008). The remaining 82% of the studies were more recently published within the past 10 years (2009-2019). These 17 articles validated 24 individual outcome measures in upper-middle income countries for use in stroke rehabilitation.

Table 4.1- Overview of studies from Upper-middle income countries

Author (Country)	Aims	Sample size	Age	Gender	Diagnosis	Study setting	Outcome measure	Outcome	ICF domain
Hale et al., 1998 (South Africa)	Examine the inter-rater reliability and concurrent validity of the Soweto Stroke Questionnaire.	n=54 Re-test testing: n=19	Not reported	Not reported	Not reported 3 months post discharged from hospital.	Participants interviewed in their home environment (various forms of housing).	Soweto Stroke Questionnaire (SSQ)	Complications post stroke; Functional ability	All domains
Oveisgharan et al., 2006 (Iran)	Translate the Barthel Index and make its Persian translated form valid and reliable.	n=459	Mean: 68.11 years (SD 11.59); Range: 22-96	Male: 243 (52.9%)	Not reported	Telephone interviews of sample from Isfahan Cardiovascular Research Center's Stroke Registry (WHO unit). In/out-patient type not specified.	Barthel Index – Persian version	Task performance; Function - ADLs Mobility	Activity limitation
Ansari et al., 2008 (Iran)	Determine the inter-rater and intrarater reliability of the Modified Modified Ashworth Scale (MMAS) in knee extensor post-stroke spasticity	n=15	Mean: 67 years Range: 62-75	Male: 7 Female: 8	All ischaemic strokes; Right hemiplegia: 9 Left hemiplegia: 6; Onset of stroke: mean 14.13 months (SD 12.77; range 1-46)	Study setting and in/out-patient type not specified.	Modified Modified Ashworth Scale (MMAS)	Muscle tone	Body structure & function
Ansari et al., 2009 (Iran)	Investigate the inter-rater reliability of the MMAS in the assessment of elbow flexor spasticity in adult patients with post-stroke hemiplegia	n=21	Mean: 60 years	Male: 16 Female:5	Right hemiplegia: 11; Left hemiplegia: 10; Onset of stroke: mean 11 months (range 5-18.5)	Physiotherapy clinic for stroke. Faculty of Rehabilitation, Tehran University of Medical Sciences. Outpatients.	Modified Modified Ashworth Scale (MMAS)	Muscle tone	Body structure & function
Kurtais et al., 2009 (Turkey)	To investigate the psychometric properties of the Rivermead Motor Assessment	n=107	Mean: 62.4 years (SD 12.8) Range: 28–85 Median: 65 years	Male: 60%	Ischaemic: 79% Haemorrhagic: 21%; Right hemiplegia:48% Onset of stroke: median 2 months (mean 5.6; SD 11.2; range 0.5–78 months)	The study was conducted in the rehabilitation unit of a university hospital. Inpatients	Rivermead Motor Assessment (RMA)	Gross function, leg and trunk, and arm	Activity limitation/ Disability

ADL = Activities of daily life; ICF= International Classification of Functioning Disability and Health; n= = sum of; SD= Standard deviation; WHO= World Health Organisation

Table 4.1 (Continued) Overview of studies from Upper-middle income countries

Author (Country)	Aims	Sample size	Age	Gender	Diagnosis	Study setting	Outcome measure	Outcome	ICF domain
Fallahpour et al., 2011 (Iran)	Evaluate the psychometric properties of the Persian version of the Impact on Participation and Autonomy questionnaire (IPA-P)	n=102	Mean: 58.3 years (SD 11.9) Range: 27-75	Male: 60 Female: 42	Ischaemic stroke: 88 (86.3%); Haemorrhagic stroke: 14 (13.7%); Onset of stroke: mean 17.7 months (SD 10.01; Range 5-36)	Two neurological wards at two university hospitals and two university rehabilitation clinics In and outpatients	Impact on Participation and Autonomy - Persian version (IPA-P)	Self-perceived participation, autonomy, problems in participation and autonomy	Participa-tion restriction
Faria et al., 2011 (Brazil)	Assess stroke subject's performance as well as the intra- and inter-rater reliability, measurement errors and the minimal detectable changes of the listed tests	n=16	Mean: 52.0 years (SD 17.1) Range: 26-81	Male: 11 (68.8%) Female: 5 (31.2%)	Ischaemic stroke: 10 (62.5%); Haemorrhagic stroke: 6 (37.5%); Onset of stroke: mean 4.9 years (SD 4.5; range 1-12.9)	Research laboratory at a university Outpatients	Comfortable and Maximal gait speed Ascending stairs cadence Descending stairs cadence 'Timed Up and Go'	Gait	Activity limitation
Kaya et al., 2011 (Turkey)	Investigate inter-rater agreement of two physicians assessing post stroke elbow flexor spasticity for both MAS and MMAS.	n= 64	Mean: 60.5 years (SD 11.9)	Male: 41 (64.1%); Female: 23 (35.9%)	Right hemiplegia: 30 (46.9%%); Left hemiplegia: 34 (53.1%); Ischemic: 57 (89.1%); Haemorrhagic: 7 (10.9%); Onset of stroke: mean 15.7 weeks (SD 10.2; range 2-28);	Physical Therapy & Rehabilitation Dept, Izmir Bozyaka Training & Research Hospital, Izmir, Turkey In/out-patient type not specified.	Modified Modified Ashworth Scale (MMAS) Modified Ashworth Scale (MAS)	Muscle tone	Body structure and function
Ansari et al., 2012 (Iran)	Develop and evaluate the interrater and intrarater reliability of the Persian version of the MMAS in post stroke subjects.	n=30;	Mean: 52.3 years (SD 13.5; range 25-81)	Male: 47% female: 53%	Right hemiplegia: 14; Left hemiplegia: 16; Onset of stroke: mean 20 months (SD 19.5; range 3-96)	Tabassom Centre of Stroke and Rehabilitation. Tehran, Iran In/out-patient type not specified.	Persian Modified Modified Ashworth Scale (P-MMAS)	Muscle tone	Body structure and function

ICF= International Classification of Functioning Disability and Health; n= = sum of; SD= Standard deviation; WHO= World Health Organisation

Table 4.1 (Continued) Overview of studies from Upper-middle income countries

Author (Country)	Aims	Sample size	Age	Gender	Diagnosis	Study setting	Outcome measure	Outcome	ICF domain
Hiengkaew et al., 2012 (Thailand)	Determine test-retest reliability and absolute & relative minimal detectable changes at the 95% confidence level of listed tests in individuals with stroke with differences in ankle plantar flexor tone.	n= 61	Mean: 63.5 years (SD 10.0)	Male: 43 (70%); female: 18 (30%)	Right hemiplegia: 28 (46%); Left hemiplegia: 33 (54%); Ischemic: 38 (62%), Haemorrhagic: 23 (38%). Onset of stroke: mean 40.2 months (SD 34.3; range 6-145);	Outpatient physical therapy clinics.	Berg Balance Scale Fugl-Meyer Assessment Scale Timed "Up & Go" Test Gait speeds 2-Minute Walk Test	Balance Lower limb function Gait	Activity limitation
Chingsongkram et al., 2014 (Thailand)	Assess the reliability and convergent validity of the BESTest in patients with subacute stroke and to determine whether the BESTest could be used to identify patients with low and high functional ability, as classified with the Fugl-Meyer Assessment motor subscale.	Reliability testing: n=12	Mean: 58.24 years	Male: 8 Female: 4	Not reported	Prasart Neurological Institute (stroke rehabilitation centre) In/out-patient type not specified.	Balance Evaluation Systems Test (BESTest)	Balance	Activity limitation
		Validity testing: n=70	Mean: 57.01 years (SD 12.23)	Male: 32 Female: 38	Ischemic stroke: 54; Haemorrhagic: 16 Onset of stroke: mean 1.11 months (SD 2.00).				
Li et al., 2014 (China)	Investigate the inter-rater and intra-rater reliability of the TSS and to analyse the relationships between TSS and MAS and MTS.	n=71	Mean: 62.3 years (SD 15.01)	Male: 50 Female: 21	Ischemic: 55; Haemorrhagic: 16 Onset of stroke: mean 14.8 months (SD 26.03)	Inpatients at a rehabilitation hospital	Triple Spasticity Scale (TSS)	Spasticity	Body structure and function
Sahathevan et al., 2014 (Malaysia)	Translate the MoCA into Bahasa Malaysia, and to determine the validity of the translated version in a bilingual Malaysian stroke population.	n=40	Mean: 57.2 (SD 10.3)	Male: 27 (68%)	Haemorrhagic: 19 (48%) Onset of stroke: mean 330 days (range 164-581)	Physiotherapy clinic of the University Kebangsaan Malaysia Medical Centre Outpatients	Montreal Cognitive Assessment (MoCA)	Cognitive function	Body structure and function
Yaliman et al., 2014 (Turkey)	Determine the interrater and intrarater reliability of the Wisconsin Gait Scale	n=19	Mean: 59.4 years (SD 9.12) Range 19-67)	Male: 14 Female: 5	Ischemic: 13; Haemorrhagic: 6; Right hemiplegia: 7.	Physical Medicine and Rehabilitation Department, Medical Faculty of Istanbul University Hospital Inpatients	Wisconsin Gait Scale (WSG)	Gait	Activity limitation

ICF= International Classification of Functioning Disability and Health; n= = sum of; SD= Standard deviation; WHO= World Health Organisation

Table 4.1 (Continued) Overview of studies from Upper-middle income countries

Author (Country)	Aims	Sample size	Age	Gender	Diagnosis	Study setting	Outcome measure	Outcome	ICF domain
Naghdi et al., 2016 (Iran)	Develop a Persian version of the FIM, and to determine its reliability and validity.	n=40	Mean: 60 years (SD 14.9) Range 22-82.	Male: 25	Ischemic: 34; Haemorrhagic: 6; Left hemiplegia: 23; Right hemiplegia: 17 Onset of stroke: mean 21 months (SD 23; range2-120)	Tabassom Center of Stroke Rehabilitation in Tehran, Iran Outpatients	Persian version of Functional Independence Measure (FIM-P)	Performance of ADLs	Activity limitation
Ostrofsky et al., 2016 (South Africa)	Assess the reliability & validity of South African dysphagia screening tool in adults presenting with stroke.	n=63/62	Not reported	Not reported	Not reported	Three government hospitals in South Africa Inpatients	The South African dysphagia screening tool	Swallow	Body structure and function
Barbosa et al., 2019 (Colombia)	Translation and cross-cultural adaptation of the Fugl-Meyer assessment (Motor domain) in Colombian Spanish.	n=10	Mean: 55 years Range: 21-83	Male: 4 Female: 6	Ischemic stroke: 8 Haemorrhagic stroke: 2	Central Military Hospital of Colombia. Inpatients	Colombian Spanish version of the Fugl-Meyer assessment (CS-FMA)	Sensorimotor function	Body structure and function

ADL = Activities of daily life; ICF= International Classification of Functioning Disability and Health; n= = sum of; SD= Standard deviation; WHO= World Health Organisation

Table 4.2 describes the four studies originating from lower-middle income countries (colour-coded pink) which were published during 2009-2019 with one study published 10 years ago, and the remaining three (75%) published more recently within the last five years (2016-2019). In these articles, five outcome measures were validated for use in lower-middle income countries.

Table 4.2 - Overview of studies from Lower-middle income countries.

Author (Country)	Aims	Sample size	Age	Gender	Diagnosis	Study setting	Outcome measure	Outcome	ICF domain
Hamzat et al., 2009 (Nigeria)	Evaluate the validity of a Yoruba translated version of the London Handicap Scale (LHS).	n=20	Mean: 55.7 years (SD 13.4)	Male: 16; Female: 4	Right hemiplegia: 12 (60%); Left hemiplegia: 8 (40%). Onset of stroke not reported.	Physiotherapy clinics of four hospitals (3 tertiary and 1 secondary) in south-west Nigeria; Outpatients	London Handicap Scale (LHS)	Mobility, occupation, physical independence, social integration, and economic self sufficiency	Activity limitation & part-icipation restriction

Table 4.3 (Continued) Overview of studies from Lower-middle income countries.

Author (Country)	Aims	Sample size	Age	Gender	Diagnosis	Study setting	Outcome measure	Outcome	ICF domain
Osundiya et al., 2016 (Nigeria)	Investigate the responsiveness of the Ibadan Stroke-Specific Pain Scale in evaluating post-stroke pain.	n=56	Not reported	Not reported	Left hemiplegia: 68%; Right hemiplegia 32%; Pain duration: 6-12 weeks	Physiotherapy clinic of the University College Hospital Outpatients	Ibadan Stroke-Specific Pain Scale (IbSSPS)	Pain	Body structure and function
Diwan et al., 2018 (India)	Validate the Gujarati translation of SIS version 3.0 and SIS 16	Validity testing: n= 26	Mean: 55.23 years (SD 12.47 Range: 35-74	Male: 5 (19.23%); Female: 21 (80.76%)	Ischemic stroke: 22 (84.61%); Haemorrhagic stroke: 4 (15.38%) Onset of stroke: mean 28.5 months (SD 32.53).	Adult neuro-rehabilitation department of SBB college of Physiotherapy Ahmedabad, Gujarat, India Outpatients	Stroke impact Scale 3.0 (SIS 3.0) Stroke Impact scale 16 (SIS 16)	Strength, Memory, Emotion, Communication, ADL, Mobility, Hand function, Social Participation	All domains
Gelisanga et al., 2019 (Philippines)	Examine the interrater and intrarater reliability, and concurrent validity of the UMCT-KE and UMCT-KF, and associations with walking ability in adults with subacute stroke.	n= 50	Mean: 51 years (SD 12)	Male: 30 (60%) Female: 20 (40%)	Type of stroke not specified; Onset of stroke: mean 68 days (SD 48)	Rehabilitation dept in a public teaching hospital in Manila, Philippines. In/out-patient type not specified.	Upright Motor Control Test Knee Extension (UMCT-KE) Upright Motor Control Test Knee flexion (UMCT-KF)	Voluntary lower limb control in the standing position.	Activity limitation

ADL = Activities of daily life; ICF= International Classification of Functioning Disability and Health; n= = sum of; SD= Standard deviation; WHO= World Health Organisation

Dept = Department; ICF= International Classification of Functioning Disability and Health; n= = sum of; SD= Standard deviation; WHO= World Health Organisation.

Table 4.3 below describes the three studies originating from low-income countries which were published in the span of six years ranging from 2014-2019. Only four outcome measures were validated for use in these low-income countries.

Table 4.4 - Overview of studies from low income countries.

Author	Aims	Sample size	Age	Gender	Diagnosis	Study setting	Outcome measure	Outcome	ICF domain
Sogbossi et al., 2014 (Benin)	Validate the Benin version of ABILOCO to assess locomotion ability in stroke patients.	n=230	Mean: 51.1 (SD 11.6)	Male: 64.3%	Left hemiplegia: 53.5% Onset of stroke: mean 21.9 months (SD 25.4)	9 rehabilitation centres in Benin (West Africa) In and outpatients	ABILOCO	Gait	Activity limitation
Kamwesiga et al., 2016 (Uganda)	Culturally adapt and determine the psychometric properties of the Stroke Impact Scale 3.0 in the Ugandan context on a small scale.	n=95	Mean: 52.4 years (SD 14.5); Range: 16-75	male: 47%; female: 53%	Right hemiplegia: 52 (55%); Left hemiplegia: 43 (45%); Ischemic: 49 (52%), Haemorrhagic: 4 (4%), No data: 42 (44%).	Mulago National Referral Hospital neurology ward; Mulago Hospital physiotherapy department; Stroke Rehabilitation Centre in Kampala; and homes of people with stroke. In and outpatients	Stroke Impact Scale 3.0 Uganda version (SIS 3.0)	Strength; Hand function; ADLs & IADLs; Mobility; Communication; Emotion; Memory and thinking; and Participation.	All domains
Niama Nata et al 2019 (Benin)	Cross cultural validation of the ABILHAND-Stroke questionnaire for post-stroke patients living in Benin.	n=233	Mean: 54 (SD9.7) Range: 28-79	Males 149 (66.8%); Females 74 (33.2%)	Ischemic 79 (35.4%); Haemorrhagic 33 (14.8%); undefined 111 (49.8%);	Outpatient rehabilitation clinics in the Physical Medicine and Rehabilitation department of the CNHUHKM, and in the Army Hospital of Cotonou	ABILHAND-stroke	Manual ability	Activity limitation

ADL = Activities of daily life; IADLs – Instrumental activities of daily living; ICF= International Classification of Functioning Disability and Health; n= = sum of; SD= Standard deviation; WHO= World Health Organisation

4.4 Appraisal of methodological procedures of studies

Table 4.4 below tabulates the results of the critical appraisal of all included studies. All of the studies made use of appropriate statistical tests to obtain their results and any withdrawals from the studies were clearly reported. Only three studies (Ansari et al., 2009; Faria et al., 2011; Barbosa et al., 2019) met all the criteria of the critical appraisal tool by Brink and Louw (2011). Criteria items 4, 5, 6, 12 and 13 regarding rater blinding, random testing order, withdrawals and statistical methods were mostly adhered to. Overall, the studies reported good descriptions of the index test but often failed to give an adequate explanation of the reference standards.

Table 4.5 - Evaluation of methodological procedures using the CAT by Brink & Louw (2011)

Author	Item 1: Sample described	Item 2: Rater qualified	Item 3: Reference standard explained	Item 4: Inter-rater blinding	Item 5: Intra-rater blinding	Item 6: Random test order	Item 7: Test period	Item 8: Retest period (index)	Item 9: Independent ref std & index test	Item 10: Index test procedures detailed	Item 11: Ref std test procedures detailed	Item 12: With-drawals	Item 13: Statistical methods used
Hale et al., 1998	No	Yes	Yes	Yes	N/A	Yes	Yes	No	No	No	No	Yes	Yes
Oveisgharan et al., 2006	No	No	Yes	Yes	N/A	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Ansari et al., 2008	Yes	Yes	N/A	Yes	No	Yes	N/A	Yes	N/A	Yes	N/A	Yes	Yes
Ansari et al., 2009	Yes	Yes	N/A	Yes	Yes	Yes	N/A	Yes	N/A	Yes	N/A	Yes	Yes
Kurtais et al., 2009	Yes	Yes	Yes	N/A	N/A	N/A	Yes	No	Yes	Yes	No	Yes	Yes
Fallahpour et al., 2011	Yes	Yes*	Yes	N/A	N/A	N/A	N/A	N/A	No	Yes	Yes	Yes	Yes
Faria et al., 2011	Yes	Yes	N/A	Yes	Yes	Yes	N/A	Yes	N/A	Yes	N/A	Yes	Yes
Kaya et al., 2011	Yes	No	N/A	Yes	Yes	Yes	N/A	Yes	N/A	Yes	N/A	Yes	Yes
Ansari et al., 2012	Yes	Yes	N/A	Yes	Yes	Yes	N/A	No	N/A	Yes	N/A	Yes	Yes
Hiengkaew et al., 2012	Yes	Yes	N/A	N/A	N/A	N/A	N/A	No	N/A	Yes	N/A	Yes	Yes
Chingsongkram et al., 2014	Yes	Yes	Yes	Yes	Yes	N/A	N/A	Yes	Yes	No	No	Yes	Yes
Li et al., 2014	Yes	Yes	N/A	Yes	Yes	No	N/A	No	Yes	Yes	Yes	Yes	Yes
Sahathevan et al., 2014	Yes	No	Yes	N/A	N/A	N/A	Yes	N/A	Yes	Yes	No	Yes	Yes
Yaliman et al., 2014	Yes	Yes	N/A	No	No	No	N/A	Yes	N/A	Yes	N/A	Yes	Yes
Naghdi et al., 2016	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Ostrofsky et al., 2016	No	No	No	N/A	N/A	N/A	Yes	No	Yes	Yes	No	Yes	Yes
Barbosa et al., 2019	Yes	Yes	N/A	Yes	Yes	Yes	N/A	Yes	N/A	Yes	N/A	Yes	Yes
Hamzat et al., 2009	Yes	Yes*	Yes	N/A	N/A	N/A	Yes	N/A	Yes	No	No	Yes	Yes
Osundiya et al., 2016	Yes	No	N/A	N/A	N/A	N/A	N/A	Yes	N/A	Yes	N/A	Yes	Yes
Diwan et al., 2018	Yes	Yes	Yes	N/A	N/A	N/A	No	N/A	Yes	No	No	Yes	Yes
Gelisanga et al., 2019	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sogbossi et al., 2014	Yes	Yes	Yes	N/A	N/A	N/A	No	Yes	Yes	No	No	Yes	Yes
Kamwesiga et al., 2016	Yes	Yes	Yes	N/A	N/A	N/A	No	N/A	No	No	No	Yes	Yes
Niama Natta et al., 2019	Yes	Yes*	Yes	N/A	N/A	N/A	No	No	Yes	No	No	Yes	Yes

Yes= Adhered to critical appraisal criteria; *= participant rater; No= not enough information or did not meet criteria; N/A= criteria not applicable to testing methods utilized; Ref Std = Reference standard

4.5 Appraisal of cross-cultural adaptation and validation procedures

Table 4.5. displays the results of the appraisal of cross-cultural adaptation as reported in the studies. The evaluation considered the extent to which the methods applied in these studies followed the guidelines as described by Beaton et al., (2011). Eleven studies were appraised. Five studies did not provide enough information regarding its translation and cross-cultural adaptation process in order to evaluate its procedures (Oveisgharan et al., 2006; Sahathevan et al., 2014; Sogbossi et al., 2014; Diwan et al., 2018; Niama Natta et al., 2019). None of the remaining six studies' pre-testing sample sizes were within the required range of 30-40 participants. The translators used in most of the studies were poorly described. Reporting of the back-translation procedures was also poorly described.

Table 4.6 - Evaluation of the cross-cultural adaptation process of included studies

Stages of CCA	Stage 1 (Translation)		Stage 2 (Synthesis)		Stage 3 (Back translation)			Stage 4 (Expert committee review)				Stage 5 (Pre-testing)			Stage 6 (through-out stages)
	Two trans-lations T1 & T2	Informed and un-informed trans-lators	Synthesize T1 & T2 into T12	Resolve any dis-crepancies with translators' reports	Two English first language translators naïve to outcome measurement	Work from T12 version	Create 2 back trans-lations BT1 & BT2	Re-view all reports	Member type	Reach consensus on dis-crepancies	Produce Pre-final version	n= 30-40	Complete questionnaire	Probe to get under-standing of item	Appraisal of reports by developer committee
Oveisgharan et al., 2006	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Fallapour et al., 2011	Yes	No	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes
Ansari et al., 2012	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	No	No
Sahathevan et al., 2011	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Naghdi et al., 2016	Yes	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes
Barbosa et al., 2019	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	No	No	Yes
Hamzat et al., 2009	No	No	No	No	Yes	Yes	No	No	No	No	Yes	No	No	No	No
Diwan et al., 2018	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Sogbossi et al., 2014	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Kamwesiga et al., 2016	No	Yes	No	No	No	No	No	Yes	Yes	No	Yes	No	No	No	Yes
Niama Natta et al., 2019	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No

Yes= Adhered to critical appraisal criteria; No= not enough information or did not meet criteria; N/A= criteria not applicable to testing methods utilized; T1 = translation 1; T2= translation 2;T12= combined translation 1&2; n= = sum of;

Tables 4.6 to 4.11. present the findings of the appraisal of the psychometric testing using the COSMIN methodology manual (Mokkink et al., 2018

Table 4.7 - Evaluation of studies investigating reliability of an outcome measure from upper middle-income countries

Study	Outcome measure	Measurement property	Psychometric test score	Interpretation
Ansari et al., 2012	Persian Modified Modified Ashworth Scale (P-MMAS)	Inter-rater reliability	k(w) = 0.91 and 0.90	Sufficient
		Intrarater reliability	k(w) = 0.81 and 0.83	Sufficient
Ansari et al., 2009	Modified Modified Ashworth Scale (MMAS)	Inter-rater reliability	k(w) = 0.81	Sufficient
Ansari et al., 2008	Modified Modified Ashworth Scale (MMAS)	Inter-rater reliability	k = 0.72	Sufficient
			Kendall tau-b = 0.87	Sufficient
		Intrarater reliability	k = 0.82	Sufficient
			Kendall tau-b = 0.92	Sufficient
Chinsongkram et al., 2014	Balance Evaluation Systems Test (BESTest)	Inter-rater reliability	Total scores: ICC=0.99	Sufficient
			Subsections: ICC range from 0.87 to 0.98	Sufficient
		Intrarater reliability	Total scores: ICC=0.99	Sufficient
			Subsections: ICC range from 0.95 to 0.99	Sufficient
Fallahpour et al., 2011	Impact on Participation and Autonomy – Persian version (IPA-P)	Scale reliability	Scale reliability= 0.92	Sufficient
Faria et al., 2011	Comfortable gait speed (mean of three trials)	Inter-rater reliability	ICC=0.97	Sufficient
		Intrarater reliability	ICC= 0.95	Sufficient
	Maximal gait speed	Inter-rater reliability	ICC=0.97	Sufficient
		Intrarater reliability	ICC=0.92	Sufficient
	Comfortable ascending stairs cadence	Inter-rater reliability	ICC=0.97	Sufficient
		Intrarater reliability	ICC=0.95	Sufficient
	Maximal ascending stairs cadence	Inter-rater reliability	ICC=0.94	Sufficient
		Intrarater reliability	ICC=0.95	Sufficient
	Comfortable descending stairs cadence	Inter-rater reliability	ICC=0.96	Sufficient
		Intrarater reliability	ICC=0.93	Sufficient
	Maximal descending stairs cadence	Inter-rater reliability	ICC=0.97	Sufficient
		Intrarater reliability	ICC=0.96	Sufficient
	'Timed Up and Go' test	Inter-rater reliability	ICC=0.96	Sufficient
		Intrarater reliability	ICC=0.85	Sufficient

Legend: AUC= Area under curve; BBS= Berg Balance Scale; CB&M= Community Balance and Mobility Scale; DIF= Differential item functioning; FAC= Functional Ambulation Classification; FIM= Functional Independence Measure; ICC= intraclass correlation coefficient; IRT= Item response theory; k(w)= weighted kappa; MTS= Modified Tardieu Scale; MRS= Modified Rankin Scale; PASS= Postural Assessment Scale for Stroke; SIS= Stroke Impact Scale; 6MWT= 6 minute walk test; 10MWT= 10m walk test;

Table 4.6. (Continued) Evaluation of studies investigating reliability of an outcome measure from upper middle-income countries

Study	Outcome measure	Measurement property	Psychometric test score	Interpretation
Hiengkaew et al., 2012	Berg Balance Scale	Test-retest reliability	ICC= 0.95	Sufficient
		Measurement error	SEM= 1.68 (MDC95=4.66)	Sufficient
	Fugl-Meyer Assessment Scale	Test-retest reliability	ICC= 0.94	Sufficient
		Measurement error	SEM= 1.29 (MDC95=3.57)	Sufficient
	Timed "Up & Go" Test	Test-retest reliability	ICC= 0.97	Sufficient
		Measurement error	SEM= 2.83 (MDC95=7.84)	Sufficient
	Comfortable gait speeds	Test-retest reliability	ICC= 0.96	Sufficient
		Measurement error	SEM= 0.06 (MDC95=0.18)	Sufficient
	Fast gait speed	Test-retest reliability	ICC= 0.98	Sufficient
		Measurement error	SEM= 0.05 (MDC95=0.13)	Sufficient
	2-Minute Walk Test	Test-retest reliability	ICC= 0.98	Sufficient
		Measurement error	SEM= 4.83 (MDC95=13.4)	Sufficient
Kaya et al., 2011	Modified Modified Ashworth Scale (MMAS)	Inter-rater agreement	k(w)= 0.892	Sufficient
	Modified Ashworth Scale (MAS)	Inter-rater agreement	k(w)= 0.868	Sufficient
Kurtais et al., 2009	Rivermead Motor Assessment (RMA)	Internal consistency	a = 0.88-0.95	Sufficient
Li et al., 2014	Triple Spasticity Scale	Test-retest reliability	ICC = 0.905~0.918	Sufficient
		Inter-rater reliability	ICC = 0.778~0.885	Sufficient
Naghdi et al., 2016	Persian version of Functional Independence Measure (P-Fim)	Internal consistency	Cronbach's alpha = 0.70 to 0.96.	Sufficient
		Inter-rater reliability	ICC= 0.88–0.98.	Sufficient
Ostrofsky et al., 2016	The South African dysphagia screening tool (SADS)	Rater agreement	Cohen's kappa= 0.51-0.60.	Insufficient
Oveisgharan et al., 2006	Barthel Index (BI)	Internal consistency	a= 0.9354.	Sufficient
		Test-retest	r= 0.989.	Sufficient
		Rater agreement	W= 0.994	Sufficient
Yaliman et al., 2014	Wisconsin Gait Scale (WGS)	Internal consistency	a= 0.91-0.94	Sufficient
		Inter-rater reliability	Total score: ICC= 0.91 – 0.96	Sufficient
		Item reliability	ICC= 0.816-1.0	Sufficient

Legend: a = alpha; r= Rho; AUC= Area under curve; BBS= Berg Balance Scale; CB&M= Community Balance and Mobility Scale; DIF= Differential item functioning; FAC= Functional Ambulation Classification; FIM= Functional Independence Measure; IRT= Item response theory; k(w) = weighted kappa; MTS= Modified Tardieu Scale; MRS= Modified Rankin Scale; PASS= Postural Assessment Scale for Stroke; SEM = Standard error of measurement; SIS= Stroke Impact Scale; 6MWT= 6 minute walk test; 10MWT= 10m walk test;

Table 4.8 - Evaluation of studies investigating reliability of an outcome measure from lower middle-income countries

Study	Outcome measure	Measurement property	Psychometric test score	Interpretation
Gelisanga et al., 2019	Upright Motor Control Test Knee Extension (UMCT-KE)	Inter-rater reliability	W=0.76–0.79	Sufficient
	Upright Motor Control Test Knee flexion (UMCT-KF)	Intrarater reliability	W=0.77-0.78	Sufficient

Legend: AUC= Area under curve; BBS= Berg Balance Scale; CB&M= Community Balance and Mobility Scale; DIF= Differential item functioning; FAC= Functional Ambulation Classification; FIM= Functional Independence Measure; IRT= Item response theory; MTS= Modified Tardieu Scale; MRS= Modified Rankin Scale; PASS= Postural Assessment Scale for Stroke; SIS= Stroke Impact Scale; W= co-efficient of concordance 6MWT= 6 minute walk test; 10MWT= 10m walk test;

Table 4.9 - Evaluation of studies investigating reliability of an outcome measure from low income countries

Study	Outcome measure	Measurement property	Psychometric test score	Interpretation
Kamwesiga et al., 2016	Stroke Impact Scale 3.0 Uganda version (SIS 3.0)	Internal consistency	a = 0.90-0.96 all domains (except memory 0.75)	Sufficient
		Scale reliability	Ranged from 0.75-0.96	Sufficient
Sogbossi et al., 2014	ABILOCO (ABILOCO)	Test-retest reliability	ICC= 0.95 for item difficulty and 0.93 for subject measures.	Sufficient
		Scale reliability	0.93	Sufficient

Legend: a= alpha; AUC= Area under curve; BBS= Berg Balance Scale; CB&M= Community Balance and Mobility Scale; DIF= Differential item functioning; FAC= Functional Ambulation Classification; FIM= Functional Independence Measure; ICC= intraclass correlation coefficient; IRT= Item response theory; MTS= Modified Tardieu Scale; MRS= Modified Rankin Scale; PASS= Postural Assessment Scale for Stroke; SIS= Stroke Impact Scale; 6MWT= 6 minute walk test; 10MWT= 10m walk test;

Sufficient inter-rater and intrarater reliability was found in all applicable outcome measures. These outcome measures included the Modified Modified Ashworth scale; BESTest; gait speeds, stair ascend/descend cadence, Timed ‘up and go’ test; Upright Motor Control Test Knee Extension & Flexion; Triple Spasticity scale; Persian version of the Functional Independence measure and the Wisconsin gait scale. The Modified Modified Ashworth scale was assessed in the Iranian sample with post-stroke knee extensor and elbow flexor spasticity (Ansari et al., 2008; Ansari et al., 2009). The Persian version of the Modified Modified Ashworth Scale was also assessed in participants with elbow flexor spasticity by Ansari et al., (2012). The BESTest was evaluated in participants from Thailand and also showed sufficient correlation with the Berg Balance Scale, the Postural Assessment Scale for Stroke, the Community Balance and Mobility Scale and the Mini-BESTest (Chinsongkram et al., 2014). Faria et al., (2011) established favourable reliability in mobility scales within a Brazilian sample population.

Table 4.10 - Evaluation of studies investigating validity of an outcome measure from upper middle-income countries

Study	Outcome measure	Measurement property	Psychometric test score	Interpretation
Chinsongkram et al., 2014	Balance Evaluation Systems Test (BESTest)	Criterion validity	BBS: r=0.96	Sufficient
			PASS: r=0.96	Sufficient
			CB&M: r=0.91	Sufficient
			Mini BESTest: r=0.96	Sufficient
		Discriminate validity	AUC=0.85	Sufficient
Fallahpour et al., 2011	Impact on Participation and Autonomy – Persian version (IPA-P)	Criterion validity	Two-tailed analysis: r>0.60	Insufficient
			Unidimensionality: supported	Indeterminate
			Scale reliability: 0.92	Sufficient
			Goodness of fit: not acceptable	Indeterminate
Hale et al., 1998	The Soweto Stroke Questionnaire (SSQ)	Criterion validity	Barthel Index: r=0.948	Sufficient
Kurtais et al., 2009	Rivermead Motor Assessment (RMA)	Structural validity	Mokken scale analysis: 0.723 & 0.927	Insufficient
			Residual correlations= 0.83 to -1.22	Insufficient
			Scale reliability: 0.95	Sufficient
			IRT: 0.27 to 6.65	Sufficient
Li et al., 2014	Triple Spasticity Scale	Criterion validity	With MAS in both elbow flexors and plantar flexors: r = 0.840-0.946	Sufficient
			With MTS, in both elbow flexors and plantar flexors: r = 0.715-0.795	Sufficient
Naghdi et al., 2016	Persian-Functional Independence Measure (P-Fim)	Criterion validity	With the Persian Barthel Index: r = 0.95	Sufficient
Oveisgharan et al., 2006	Barthel Index (BI)	Criterion validity	With MRS: r= –0.912 (p<0.001).	Insufficient
Sahathevan et al., 2014	Montreal Cognitive Assessment-BM (MoCA)	Criterion validity	With English MoCA: ICC= 0.81	Sufficient

Legend: AUC= Area under curve; BBS= Berg Balance Scale; CB&M= Community Balance and Mobility Scale; DIF= Differential item functioning; FAC= Functional Ambulation Classification; FIM= Functional Independence Measure; IRT= Item response theory; MTS= Modified Tardieu Scale; MRS= Modified Rankin Scale; PASS= Postural Assessment Scale for Stroke; SIS= Stroke Impact Scale; 6MWT= 6 minute walk test; 10MWT= 10m walk test;

Concurrent validity tested in all studies from upper middle-income countries with the three impairment outcome measures showing sufficient validity. The validity of outcome measures covering the other ICF domains varied.

Table 4.11 - Results of studies investigating validity of an outcome measure from lower middle-income countries

Study	Outcome measure	Measurement property	Psychometric test score		Interpretation
Diwan et al., 2018	Gujarati translated version of the Stroke Impact Scale (SIS)	Criterion validity	SIS Strength with NHISS motor: r= -0.254		Insufficient
			SIS memory with MMSE: r=0.384		Insufficient
			SIS communication with NIHSS & FIM communication: r= 0.559		Insufficient
			SIS ADL with FIM Motor (Locomotion & Mobility: r= 0.323		Insufficient
			SIS mobility with:	FIM Motor: r= 0.537	Insufficient
				STREAM Basic Mobility: r= 0.646	Insufficient
				Gait speed: r 0.273	Insufficient
			SIS hand function with STREAM Upper Extremity: r= 0.490		Insufficient
			SIS Participation with FIM Social and Cognition: r= 0.292		Insufficient
			SIS Physical with:	FIM Motor: r= 0.291	Insufficient
				STREAM Basic Mobility: r= 0.263	Insufficient
Gelisanga et al., 2019	Upright Motor Control Test Knee Extension (UMCT-KE)	Criterion validity	Motricity Index leg subscale= 0.747		Sufficient
	Upright Motor Control Test Knee flexion (UMCT-KF)	Criterion validity	Motricity Index leg subscale= 0,775		Sufficient
Hamzat et al., 2009	London Handicap Scale (Yoruba-LHS)	Concurrent validity	LHS English version: r= 0.55 (p=0.002)		Insufficient

Legend: AUC= Area under curve; BBS= Berg Balance Scale; CB&M= Community Balance and Mobility Scale; DIF= Differential item functioning; FAC= Functional Ambulation Classification; FIM= Functional Independence Measure; IRT= Item response theory; MTS= Modified Tardieu Scale; MRS= Modified Rankin Scale; PASS= Postural Assessment Scale for Stroke; SIS= Stroke Impact Scale; 6MWT= 6 minute walk test; 10MWT= 10m walk test;

Concurrent validity tested in all studies from lower middle-income countries with the two impairment outcome measures showing sufficient validity. The validity of outcome measures covering the other ICF domains were insufficient.

Table 4.12 - Results of studies investigating validity of an outcome measure from low income countries

Study	Outcome measure	Measurement property	Psychometric test score	Interpretation
Kamwesiga et al., 2016	Stroke Impact Scale 3.0 Uganda version (SIS 3.0)	Structural validity	DIF: Not reported	Indeterminate
			Residuals: Not reported	Indeterminate
			Scale reliability: 0.90 to 0.75	Sufficient
			Infit MnSq: 0.47 to 1.99; z: -5.58 to 5.89	Insufficient

Legend: AUC= Area under curve; BBS= Berg Balance Scale; CB&M= Community Balance and Mobility Scale; DIF= Differential item functioning; FAC= Functional Ambulation Classification; FIM= Functional Independence Measure; MnSq = Mean square; IRT= Item response theory; MTS= Modified Tardieu Scale; MRS= Modified Rankin Scale; PASS= Postural Assessment Scale for Stroke; SIS= Stroke Impact Scale; 6MWT= 6 minute walk test; 10MWT= 10m walk test;

Table 4.11.(Continued) Results of studies investigating validity of an outcome measure from low income countries

Study	Outcome measure	Measurement property	Psychometric test score	Interpretation
Niama Natta et al., 2019	ABILHAND-stroke (ABILHAND-S)	Structural validity	DIF= 0.64 to 0.85	Insufficient
			Residual correlations= 0.93 to -1.13	Insufficient
			Scale reliability: 0.82	Sufficient
			IRT: 0.54 to 5.69	Sufficient
Sogbossi et al., 2014	ABILOCO (ABILOCO)	Structural validity	DIF analysis= Not reported	Indeterminate
			Residual correlations: 0.31—0.26	Sufficient
			Scale reliability: 0.93	Sufficient
			IRT= 0.22-3.13	Sufficient
		Criterion validity	With FAC: r= 0.86	Sufficient
			With FIM-Mobility: r= 0.87	Sufficient
			With 6MWT: r= 0.76	Sufficient
			With 10MWT: r= 0.75	Sufficient

Legend: AUC= Area under curve; BBS= Berg Balance Scale; CB&M= Community Balance and Mobility Scale; DIF= Differential item functioning; FAC= Functional Ambulation Classification; FIM= Functional Independence Measure; IRT= Item response theory; MTS= Modified Tardieu Scale; MRS= Modified Rankin Scale; PASS= Postural Assessment Scale for Stroke; SIS= Stroke Impact Scale; 6MWT= 6 minute walk test; 10MWT= 10m walk test;

Concurrent validity tested was tested in all studies from low-income countries. The three activity-limitation domain outcome measures showed varying validity.

4.6 Conclusion

The database searches produced 24 eligible studies that were included in the review. These studies attempted to validate a total of 33 outcome measures and an analysis of the quality of these studies was done. It was found that majority of these outcome measures did not fully adhere to cross-cultural adaptation processes and displayed irregularities in its methodology which resulted in low-measure properties of outcome measures. Further analysis of the quality of these outcome measures will be discussed in the following Chapter 5: Discussion.

Chapter 5: Discussion

An analysis of trends and associations arising from the results will be discuss within this section. Comparisons will be made in order to link the data extracted from the included studies with key aspects with evidence provided in the literature review.

5.1 Overview of methods

The aim of the systematic review was to search literature reporting on the validation of outcome measures used in stroke rehabilitation within low- and middle-income countries. The secondary objectives were to describe the outcome measures specifically the populations and countries within which these outcome measures were validated for stroke rehabilitation. We conducted an extensive search of databases and the detailed findings were reported in the previous chapter. This chapter presents an interpretation of these key findings as well as the implications of the methodological and cross-cultural procedures of studies included in this review.

5.2 Outcome measuses validated in lower income countries

Our review findings shows that in lower income countries, there is sparse published information about outcome measurement validation for people with stroke. Only three (13%) studies (Sogbossi et al., 2014; Kamwesiga et al., 2016; Niama Natta et al., 2019) of the 24 studies were conducted in low income countries and only five (21%) in lower-middle income countries (Hamzat et al., 2009; Diwan et al., 2018; Barbosa et al., 2019; Gelisanga et al., 2019; Osundiya et al., 2019). This lack of information may reflect the status of stroke rehabilitation which remains sub-optimal due to a lack of human and structural resources (Yan et al., 2016). It is very likely that therapists in low income settings are not aware and therefore are not routinely using outcome measurement in practice (Demers et al., 2019). Furthermore, low income settings invest very little or no resources for rehabilitation research (Parnes et al., 2009). These factors may imply why there is very little research and consequently few publications from these settings. Similar findings were reported in other stroke related reviews. For example, a scoping review by Chimatiro et al., (2019) showed that seven studies highlighted the importance of rehabilitation in stroke care, however, the interventions used remains limited in most LMIC. These findings imply that outcome measurement in people with stroke may not be done routinely, or if used, the

outcome have not been validated for their specific context. This further indicates that therapists in low-income settings cannot show impact and this further prevents the strategies and data which is needed to convince governments to invest in rehabilitation. Wherever possible, studies taking place in low income countries can obtain guidance from research conducted in higher income countries. In this review, the majority of outcome measures were validated for use in upper-middle income countries.

5.3 Validated outcome measures within the ICF framework

Most of the outcome measures assessed the body structure and function domain and activity limitation domain of the ICF. However, only one study (Hale et al., 1998) validated an outcome measure (The Soweto Stroke Questionnaire) in a rural setting. All the other studies included in our review took place in urban areas (Maleka, Stewart & Hale, 2012). Since inequality is rife in these countries, people in rural and remote areas may have different needs compared to those in urban areas (Maleka, Stewart & Hale, 2012). The performance of certain tasks may be different due to infrastructural or resource differences; for example, toilets could be located outside an informal dwelling rather than inside a brick and mortar house, which would alter the assessment of this activity in different context (Breytenbach, 2016). In addition, the ability to perform specific tasks, such as toileting, may require a higher level of functioning when performed in a rural setting such as walking outdoors over uneven terrain to reach the communal toilet a distance from the home

Thus, in order for the findings of outcome measures to be relevant, items of the outcome measure should reflect the actual task performance in a specific context. Outcome measures originally developed in different settings such as in high income countries, are often not directly applicable or appropriate for low-income countries. Therefore, it is encouraging that more psychometric testing of outcome measures are being conducted in middle- and low-income countries as stroke incidence is escalating in these populations. Future research should also consider marginalised groups such as those residing in rural areas.

Very few outcome measures in this review assesses reintegration of participants into their community over all income groups. Only 20% (Hale et al., 1998; Fallahpour et al., 2011; Kamwesiga et al., 2016; Diwan et al., 2018; Hamzat et al., 2019) of the studies' outcome measures fell within the participation restriction ICF domain. However, there has been a

steady and recent increase in the publication of studies utilizing outcome measures which incorporate re-integration (Ntsiea, 2019). This increasing field of literature is reassuring and may imply awareness and an application of outcome measurement in this domain. The rise in the development and use of participation measures could indicate a shift of focus in stroke rehabilitation from impairments towards community reintegration, and ultimately a return to productive function if possible.

5.4 Methodological quality of included studies

The methodological appraisal of the bulk of the included studies showed minimal flaws, but three studies (Ansari et al., 2009, Faria et al., 2011, Barbosa et al., 2019) met all the criteria of the Brink and Louw (2011) critical appraisal tool. The largest shortcoming in most of the studies was item 11, which requires a detailed description of the reference standard used in the study. Lack of reporting of important information in the translation and cross-cultural adaptation procedures and inadequate pre-testing sample size were other common shortfalls noted. These types of errors could be due to a lack of guidance when drawing up the study protocol. The may be a result of limited academic resources in lower income countries.

The accuracy of results obtained in a study is dependent on the quality of the methodology used in the studies. Therefore, good and sound methodological processes ensure that results are dependable and allow for appropriate conclusions which can influence the management of patients or within research. Poor to fair methodological quality of studies was also reported in a systematic review of Brazilian Portuguese patient-reported outcome measures for use in patients with cancer Albach et al., (2018).

5.5 Cross-cultural adaptation processes of included studies

The cross-cultural adaptation (CCA) procedures followed were done fairly well in four of the six studies from upper-middle income countries. Two studies (Oveisgharan et al., 2006; Sahathevan et al., 2011) did not report their CCA procedures. Overall, multiple steps in the different stages of CCA as described by Beaton et al., (2000) were not adhered to, but the Synthesis process (stage 2) and the expert committee review (stage 4) of the adapted version of outcome measures were done well in these studies from upper-middle income countries. The extent to which CCA guidelines are adhered to will determine the quality and

efficacy of the outcome measure being adapted. A poorly developed outcome measure can limit its use and restrict the information obtained (Stewart et al., 2012). The abovementioned study by Albach et al., (2018) also found inconsistent quality of the cross-cultural adaptation process throughout the included studies from Brazil.

Three of the five studies (Sogbossi et al., 2014; Diwan et al., 2018; Niama Natta et al., 2019) from lower-middle- and low-income countries did not describe its CCA protocol. Poor CCA procedures were reported in the remaining two studies (Hamzat et al., 2009, Kamwesiga et al., 2016). Overall, none of the criteria was met in the synthesis (stage 2) of the translated versions and in the pre-testing (stage 5) of the adapted version. The challenges met in lower-middle- and low-income countries across all stages of CCA could be due to shortage of academic resources such as universities and adequately trained professionals. Limited expertise in the field of CCA of outcome measures could impact adherence to the correct procedures. The use of outcome measures with shortcomings could result in items not relevant to the new setting/context. This could be due to poor understanding of items or low objectivity in scoring which can influence results of the study and may limit the generalizability of results across populations (Stewart et al., 2012; Beaton et al., (2000).

5.6 Psychometric properties of outcome measures

All aspects of reliability which were assessed in the outcome measures were found to be sufficient except for the rater agreement of The South African dysphagia screening tool (Ostrofsky et al., 2016) which was insufficient. There was good adherence to rater blinding and randomised testing order throughout the studies. Adherence to blinding and varying testing order reduced the risk of bias which may be introduced into the studies. It also aims to obtain dependable results from the outcome measure under investigation. Raters' ability to recall scores previously tested can influence the results of a study. Varying the testing order further reduces the raters' likelihood to recall previous test scores and prevents systemic bias. Impact of bias in a study could render inconclusive results and an outcome measure that is inadequate for use.

Concurrent validity was evaluated in all outcome measures and the scores varied in all three income groups. Just over half of the studies (Hale et al., 1998; Chinsongkram et al., 2014; Li et al., 2014; Sahathevan et al., 2014; Naghdi et al., 2016) assessing validity of

outcome measures in upper-middle income countries was found to be sufficient. Two of the three studies (Hamzat et al., 2009; Diwan et al., 2018) assessing validity of outcome measures in lower-middle income countries was found to be insufficient. All three outcome measures from low income countries showed fluctuating validity when compared with a reference standard. Outcome measures that have been adequately adapted for use in a different setting allow for credible results. Stewart et al., (2012) attempts to highlight the understanding of modifications to outcome measures in order to address differences in socioeconomic status, race/ethnicity, language and literacy.

5.7 Validated outcome measures groups

Eleven outcome measures were well validated for use in LMICs in the included studies. These included the ABI-LOCO; Balance Evaluation Systems Test (BESTest); Functional Independence Measure Persian version (FIM-P); Gait performance-based tests; Modified Ashworth Scale, Modified Modified Ashworth Scale and its Persian version; Montreal Cognitive Assessment Bahasa Malaysia version (BM MoCA); Stroke-Specific Pain Scale Ibadan version (Ib-SSPS); Upright Motor Control Test (UMCT); Wisconsin Gait Scale (WGS)

The ABILOCO is a validated outcome measure of functional locomotion ability tested in a low-income Beninese stroke population. Sogbossi et al., (2014) found good construct validity when compared with the FIM-mobility and the 10MWT. These results are similar to that demonstrated in a high-income Belgian sample, reporting high correlation with the FIM walking ability and walking speed as measured by the 10MWT (Caty et al., 2008). Thus, the ABILOCO can produce credible results when assessing locomotion in a post stroke population in both low- and higher-income settings. As the FIM has training and licensing implications for users, we would recommend the ABILOCO and 10MWT as user-friendly, validated and accessible alternatives for clinicians and researchers in lower income countries.

Chinsongkram et al., (2014) assessed the BESTest in participants with subacute stroke. Stringent methodological processes were followed during the testing period which included the use of videotaped assessments. Psychometric testing showed sufficient interrater and intrarater reliability and sufficient correlation with the following gold standard measures: Berg Balance Scale, Postural Assessment Scale, Community Balance and Mobility Scale,

and MiniBESTest. These reference standards tests may be used in retraining balance and mobility in post stroke rehabilitation of patients in an upper middle-income country. However, the BESTest is recommended as it provides an indication of the underlying balance impairment.

Naghdi et al., (2016) developed the FIM-P under excellent cross-cultural adaptation processes and the evaluation of the FIM-P was found to have sufficient reliability. The results are similar to that obtained by Hsueh et al., (2002) in a Taiwanese stroke sample. It is important to note that the FIM-P was developed for a specific population and will require additional adaptation prior to use in a different setting. This is to maintain credibility of results within this new context.

Gait performance was assessed by Faria et al., (2011) in a Brazilian stroke population evaluating gait speeds (comfortable and maximal) and stair ascending and descending cadence (comfortable and maximal), and the 'Timed Up and Go', and was found to be reliable. A similar test was conducted in a Swedish stroke population with similar results by Flansbjer et al., (2005). Excellent methodological processes were adhered to by Faria et al., (2011) with a seven-day retest period which did not allow for any changes in condition that could alter results and aids in reducing recall bias. Hiengkaew et al., (2012) conducted a similar study with a 5-10 day retest period and replicated testing conditions between testing periods. Hiengkaew et al., (2012) showed sufficient test-retest reliability and measurement error of the Timed "Up and Go' and gait speed tests. Thus, these gait performance-based tests are reliable and recommended for use in chronic post stroke patients from an upper-middle- and high-income countries

Four studies (Ansari et al., 2008, Ansari et al., 2009; Kaya et al., 2011; Ansari et al., 2012;) included in this review evaluated the measurement properties of the MAS and MMAS. All four studies showed good reliability results. The Persian version of the MMAS (Ansari et al., 2012) underwent a cultural adaptation processes which was adhered to in eleven of the fifteen steps (73%). The raters used in this study were not offered any training and had minimal experience using the outcome measure. It should be taken into consideration that the five-minute interval between retesting is considered short and could potentially allow for a change in condition (muscle tone affected by previous test) which could possibly limit agreement between raters. Thus, the MAS, MMAS and its Persian version is

recommended to assess post stroke elbow flexor spasticity and knee extensor spasticity in upper middle-income countries and can be used by inexperienced raters.

The MoCA was translated to Bahasa Malaysia (BM-MoCA) and its validity was assessed against the original English version and showed a strong agreement between the two versions (Sahathevan et al., 2014). A clinical review by Julayanont et al., (2012) showed that the internal consistency of the original MoCA was good. Despite limited reporting of cross-cultural adaptation procedures followed by Sahathevan et al., (2014), the BM-MoCA is recommended as an outcome measure to assess cognitive function in a stroke population from an upper middle-income setting.

Osundiya et al., (2016) used good processes to adapt the Ib-SSPS to its specific context. However, the raters used in the reliability testing were not described. Therefore, the IbSSPS is a sensitive and responsive measure which is recommended for use in stroke rehabilitation in upper middle-income countries and to allow for training of raters when using the outcome measure to ensure accurate procedures were followed.

The Knee Flexion and Knee Extension subscales of the UMCT were assessed by Gelisanga et al., (2019). Good methods were followed making use of a two-day interval between tests and blinding of raters. The psychometric tests showed sufficient interrater and intrarater reliability of these tests. Therefore, the UMCT-KN and UMCT-KF can be recommended as reliable outcome measures for the assessment of lower limb function in persons with stroke from lower middle-income countries.

Yaliman et al., (2014) assessed the WGS in participants with subacute stroke. Stringent methodological processes were followed during the testing period which included use of videotaped assessments. This reduced the risk of bias in the study. Psychometric testing showed sufficient reliability, therefore the WGS is recommended for use as an outcome measure which can be used in gait retraining in post stroke patients in an upper middle-income country.

The following outcome measures were used in various included studies but showed flaws in their methodological processes: ABILHAND-Stroke Questionnaire (ABILHAND); Barthel Index; Fugl-Meyer Assessment (FMA); Impact on Participation and Autonomy Persian version (IPA-P); London Handicap Scale (LHS); Rivermead Motor Assessment (RMA);

Stroke Impact Scale (SIS); South African Dysphagia screening tool (SADS); Soweto Stroke Questionnaire (SSQ); and the Triple Spasticity Scale (TSS). The studies that conducted cross-cultural adaptation of these ten outcome measures did not adhere to all the processes. It is recommended that these outcome measures undergo further evaluation of their psychometric properties with greater adherence to published guidelines and more stringent procedures throughout the adaptation process.

5.8 Conclusion

According to the review findings and currently available literature, there are 33 outcome measures that have undergone validation for use in LMICs. These outcome measures span across all the domains of the ICF. However, this limited number of outcome measures is inadequate for the vast population covering LMICs. As a result, healthcare professionals and patients have limited resources in terms of access to adequate rehabilitation – and specifically – outcome measures. This can have an influence on the clinical care and management of persons with stroke in LMICs with a far-reaching impact on the potential of the patients, their families and caregivers, as well as within society.

Chapter 6: Conclusion

In this section overall conclusions will be drawn for this study. In addition, the limitations of this study will be reflected upon and any recommendations for future research will be made.

6.1 Limitations of the study

The databases searched were limited to those with access available on the University of Stellenbosch's library website during the period of data collection. This could have limited the results as some articles may have met the inclusion criteria but were published in journals within a restricted access database.

An additional limitation was placed on eligible articles with the restriction of those published in the English Language. This was due to time and financial constraints regarding the appropriate translation process of articles from a foreign language. This poses a risk of introducing a language bias into the study. However, the risk of incorrectly excluding eligible articles are reduced as English is generally perceived to be the universal language of science.

6.2 Recommendations

There are 24 studies included in this review that evaluated outcome measures in LMICs. The 33 outcome measures were evaluated within these studies and spanned across all domains of the ICF. After the evaluation of the methodological quality of the studies and an analysis of the psychometric properties of their outcome measures and correlation with reference standards, a final total of 23 outcome measures was recommended for use in LMICs. These outcome measures include:

ABILOCO; 10MWT; BESTest; Berg Balance Scale; Postural Assessment Scale; Community Balance and Mobility scale; MiniBESTest; FIM-P; Comfortable gait speed; Maximal gait speed; Comfortable ascending stairs; Maximal ascending stairs; Comfortable descending stairs; Maximal descending stairs; Timed 'Up and Go'; Modified Ashworth Scale; Modified Modified Ashworth Scale; Persian version of the Modified Ashworth Scale; Bahasa Malaysian version of the Montreal Cognitive Assessment; Ibadan version of the Stroke Specific Pain Scale; Upright Motor Control Test (Knee Flexion subscale & Knee

Extension subscale); Wisconsin Gait Scale.

It is recommended that the following outcome measures undergo further evaluation of their psychometric properties with greater adherence to published guidelines and more stringent procedures throughout the adaptation process. ABILHAND-Stroke Questionnaire (ABILHAND); Barthel Index; Fugl-Meyer Assessment (FMA); Impact on Participation and Autonomy Persian version (IPA-P); London Handicap Scale (LHS); Rivermead Motor Assessment (RMA); Stroke Impact Scale (SIS); South African Dysphagia screening tool (SADS); Soweto Stroke Questionnaire (SSQ); and the Triple Spasticity Scale (TSS).

It is also recommended that future research into adaptations to outcome measures be evaluated for use in lower income countries, due to their increasing trend in stroke prevalence.

6.3 Conclusion

The process of adapting an outcome measure for use in a different setting such as in a LMIC, or in a different sample population, can become complex. Established guidelines demonstrating the methodological and cross-cultural adaptation processes are best to use and be adhered to. This is necessary to ensure the new version or newly developed outcome measure can withstand intensive evaluation of its measurement properties. In this way, the results obtained using these outcome measures are reliable for research and can lead to more effective clinical management of patients and subsequent improved outcomes of these patients. Robust, valid and reliable stroke rehabilitation outcome measures guides assessment, clinical decision making, management planning and monitoring of effectiveness of interventions in daily clinical practice. The consistent use of robust outcome measures also allow for sound research findings and could assist in the development of future governing policies for the management of people with stroke globally.

Reference list

Agrell, B. & Dehlin, O. (2000). Mini mental state examination in geriatric stroke patients. Validity, differences between subgroups of patients, and relationships to somatic and mental variables. *Aging.* 12(6):439-444.

Albach, C.A. Wagland, R. & Hunt, K.J. (2018). Cross-cultural adaptation and measurement properties of generic and cancer-related patient-reported outcome measures (PROMs) for use with cancer patients in Brazil: a systematic review. *Quality of Life Research [*Online] 27(1):857-870*.* Available at: https://doi.org/10.1007/s11136-017-1703-5 [Accessed on 23 November 2019].

Ansari, N.N., Naghdi, S., Forogh, B., Hasson, S., Atashband, M. & Lashgari, E. (2012). Development of the Persian version of the Modified Modified Ashworth Scale: translation, adaptation, and examination of interrater and intrarater reliability in patients with poststroke elbow flexor spasticity. Disability and Rehabilitation. 34(21): 1843-1847.

Ansari, N.N., Naghdi, S., Hasson, S., Mousakhani, A., Nouriyan, A. & Omidvar, Z. (2009). Inter-rater reliability of the Modified Modified Ashworth Scale as a clinical tool in measurements of post-stroke elbow flexor spasticity. NeuroRehabilitation. 24(1):225-229.

Ansari, N.N., Naghdi, S., Younesian, P. & Shayeghan, M. (2008). Inter- and intrarater reliability of the Modified Modified Ashworth Scale in patients with knee extensor poststroke spasticity. Physiotherapy Theory and Practice. 24(3):205-213.

Barbosa, N.E., Forero, S.M., Galeano, C.P., Hernandez, E.D., Landinez, N.S., Sunnerhagen, K.S. & Murphy, M.A. (2019). Translation and cultural validation of clinical observational scales - the Fugl-Meyer assessment for post stroke sensorimotor function in Colombian Spanish. Disability and Rehabilitation. 41(19):2317-2323.

Beaton, E.D., Bombardier, C., Guillemin, F. & Ferraz, M.B. (2000). Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine*. 25(24):3186-3191.

Berg, K.O., Wood-Dauphinee, S.L., Wiliiams, J.I. & Maki, B. (1992). Measuring Balance in the Elderly: Validation of an Instrument. *Canadian Journal of Public Health*. [Online]. 83(2):S7-S11. Available at https://www.jstor.org/stable/41990843. [Accessed on 06/10/2019].

Berg, K.O., Wood-Dauphinee, S.L. & Wiliiams, J.I. (1995). The balance scale: Reliability assessment with elderly residents and patients with an acute stroke. *Scandinavian Journal of Rehabilitation Medicine* 27(1):27-36.

Bohannon, R.W. & Smith, M.B. (1987). Interrater reliability of a modified Ashworth scale of muscle spasticity. *Physical Therapy*. 67(2):206-207. Available at: doi: 10.1093/ptj/67.2.206 [Accessed on: 08 October 2019].

Breytenbach, F. (2016). Content validity of the modified Barthel Index for stroke patients in South Africa. Unpublished research report. *University of the Witwatersrand.* Available at: https://wiredspace.wits.ac.za [Accessed on 18 February 2018].

Brink, Y. & Louw, QA. (2011). Clinical instruments: reliability and validity critical appraisal. *Journal of Evaluation in Clinical Practice*, [online] 18(6): 1126-1132. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2753.2011.01707 [Accessed 22 June 2018].

Canadian Partnership for Stroke Recovery: Stroke Engine (c2019). Modified Rankin Scale. Available at: https://www.strokengine.ca/en/psycho/mrs_psycho [Accessed on 20 October 2019].

Caty, G.D., Arnould, C., Stoquart, G.G., Thonnard, J.L. & Lejeune, T.M. (2008). ABILOCO: a Rasch-built 13-item questionnaire to assess locomotion ability in stroke patients. *Archives of Physical Medicine and Rehabilitation.* 89(2):284-290. Available at https://dx.doi.10.1016/j.apmr.2007.08.155 [Accessed on 13 October 2019].

Chimatiro, G.L. & Rhoda, A.J. (2019). Scoping review of acute stroke care management and rehabilitation in low and middle-income countries. *BioMedCentral [Online] 19(789):1-15. Available at:* https://doi.org/10.1186/s12913-019-4654-4 *[Accessed on 22 November 2019].*

Chinsongkram, B., Chaikeeree, N., Saengsirisuwan, V., Horak, F.B. & Boonsinsukh, R. (2016). Responsiveness of the Balance Evaluation Systems Test (BESTest) in people with subacute stroke. *Physical Therapy*. 96(1): 1638-1647. Available at https://dx.doi.10.2522/ptj.20150621 [Accessed on 13 October 2019].

Chinsongkram, B., Chaikeeree, N., Saengsirisuwan, V., Viriyatharakij, N., Horak, F.B. & Boonsinsukh, R. (2014). Reliability and Validity of the Balance Evaluation Systems Test (BESTest) in People With Subacute Stroke. Physical Therapy. 94(11):1632-1643.

Coster, WJ. & Mancini, M. C. (2015). Recommendations for translations and cross-cultural adaptation of instrument for occupational therapy research and practice. *Revista de Terapia Ocupacional da Universidade de Sao Paulo,* [online]. 26(1):50-57. Available at: http://dx.doi.org/10.11606/issn.2238-6149.v26i1p50-7 [Accessed 04 March 2018].

Collen, F.M., Wade, D.T. & Bradshaw, C.M. (1990). Mobility after stroke: reliability of measures of impairment and disability. *International Disability Studies.* 12(1):6-9.

Cunningham, N. & Rhoda, A. (2014). Outcomes of stroke patients discharged from an in-patient facility in the Eastern Cape, South Africa: A mixed methods design. *South African Journal of Physiotherapy,* [online] 70(3): 26-31. Available at: https://sajp.co.za/index.php/sajp/article/view/265/244 [Accessed on: 04 August 2018].

Demers, M., Blanchette, A.K., Mullick, A.A., Shah, A., Woo, K., Solomon, J. & Levin, M.F. (2019). Facilitators and barriers to using neurological outcome measures in developed and developing countries. *Physiotherapy Research International.* 24(1):1-9. Available at: https://doi.org/10.1002/pri.1756 [Accessed on 27 November 2019].

Dennis, M., Mead, G., Doubal, F. & Graham, C. (2012). Determining the modified Rankin score after stroke by postal and telephone questionnaires. *Stroke.* 43(3):851-853. Available at https://dx.doi.10.1161/STROKEAHA.111.639708 [Accessed on 20 October 2019].

Dick, J.P., Guiloff, R.J., Stewart, A., Blackstock, J., Bielawska, C., Paul, E.A. & Marsden, C.D. (1984). Mini-Mental state examination in neurological patients. *Journal of Neurology, Neurosurgery and Psychiatry.* 47(1): 496-499.

Diwan, S.J., Shah, Z.R. & Joshi, P.B. (2018). Validation of Gujarati Translated Version of Stroke Impact Scale. Journal of Clinical and Diagnostic Research.12(8):13-16.

Duncan, P.W., Wallace, D., Lai, S.M., Johnson, D., Embretson, S. & Laster, L.J. (1999). The stroke impact scale version 2.0. Evaluation of reliability, validity, and sensitivity to change. *Stroke*. 30(10):2131-2140.

Fallahpour, M., Jonsson, H., Joghataei, M.T. & Kottorp, A. (2011). Impact on Participation and Autonomy (IPA): Psychometric evaluation of the Persian version to use for persons with stroke. Scandinavian Journal of Occupational Therapy. 18:59-71.

Faria, C.D.C.M., Teixeira-Salmela, L.F., Neto, M.G. & Rodrigues-de-Paula, F. (2011). Performance-based tests in subjects with stroke: outcome scores, reliability and measurement errors. Clinical Rehabilitation. 26(5):460-469.

Feigin, V.L., Lawes, C.M.M., Bennett, D.A., Barker-Collo, S.L. & Parag, V. (2009) Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. *Lancet Neurology* [Online]. 8(1):355-369. Available at: https://dx.doi.10.1016/S1474-4422(09)70025-0 [Accessed on: 07 October 2019].

Flansbjer, U., Blom, J. & Brogardh, C. (2012). The reproducibility of Berg Balance Scale and the Single-Leg Stance in chronic stroke and the relationship between the two tests. *Physical Medicine and Rehabilitation* 4(1):165-170. Available at https://dx.doi.10.1016/j.pmrj.2011.11.004 [Accessed on 13 October 2019].

Flansbjer, U.B., Holmback, A.M., Downham, D., Patten, C. & Lexell, J. (2005). Reliability of gait performance tests in men and women with hemiparesis after stroke. *Journal of Rehabilitation Medicine.* 37(2):75-82.

Folstein, M.F., Folstein, S.E. & McHugh, P.R. (1975). “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research.* 12(3): 189:198.

Gelisanga, M.A.P. & Gorgon, E.J.R. (2019). Measurement properties of the upright motor control test in adults with subacute stroke. Topics in Stroke Rehabilitation. 26(1)18-23.

Gbiri, CA., Olawale, OA. & Isaac, SO. (2014) Stroke management: Informal caregivers' burden and strains of caring for stroke survivors. *Annals of Physical and Rehabilitation Medicine*, [online]. 58(2): 98-103. Available at: https://dx.doi.org/10.1016/j.rehab.2014.09.017 [Accessed on 04 August 2018]

Gregson, J.M., Leathley, M.J., Moore, A.P., Smith, T.L., Sharma, A.K. & Watkins, C.L. (2000). Reliability of measurements of muscle tone and muscle power in stroke patients. *Age and Ageing* [Online] 29(3):223-228. Available at: https://dx.doi.10.1093/ageing/29.3.223 [Accessed on 20 October 2019].

Harwood, R.H., Rogers, A., Dickson, E. & Ebrahim, S. (1994). Measuring handicap: the London Handicap Scale, a new outcome measure for chronic disease. *Quality Health Care.* 3(1):11-16.

Hale, L.A., Eales, C.J. & Fritz, V.U. (1998) The Soweto Stroke Questionnaire. South African Journal of Physiotherapy.54(4):16-20.

Hamzat, TK. & Peters, GO. (2009) The London Handicap Scale: Validation of a Yoruba (Nigerian) version among stroke survivors. African Journal of Neurological Sciences [Online]. 28(1): 61-66.

Han, S. (2018). Everything you should know about ischemic stroke. *Healthline.* [Online]. Available at: https://www.healthline.com/health/stroke/cerebral-ischemia [Accessed on 25 November 2019].

Hiengkaew, V., Jitaree, K. & Chaiyawat, P. (2012). Minimal Detectable Changes of the Berg Balance Scale, Fugl-Meyer Assessment Scale, Timed "Up & Go" Test, Gait Speeds, and 2-Minute Walk Test in Individuals With Chronic Stroke With Different Degrees of Ankle Plantarflexor Tone. Archives Physical Medicine Rehabilitation. 93(1):1201-1207.

Hsueh, I.P., Lin, J.H., Jeng, J.S. & Hsieh, C.L. (2002). Comparison of the psychometric characteristics of the Functional Independence Measure, 5 item Barthel Index, and 10 item Barthel Index in patients with stroke. *Journal of Neurology, Neurosurgery and Psychiatry.* 73(2):188-190. Available at: https://dx.doi.10.1136/jnnp.73.2.188 [Accessed on 20 October 2019].

Inglis, G., Faure, M. & Frieg, A. (2008). The awareness and use of outcome measures by South African physiotherapists. *South African Journal of Physiotherapy* [online]. 64(2): 5-11. Available at: https://sajp.co.za/index.php/sajp/article/view/102/99 [Accessed 18 June 2018].

Jette, DU., Halbert, J., Iverson, C., Miceli, E. & Shah, P. (2009). Use of standardized outcome measures in physical therapist practice: perception and applications. *Physical Therapy* [online], 89(2): 125-35. https://dx.doi.10.2522/ptj.20080234 [Accessed 21 June 2018].

Johnson, W., Onuma, O., Owolabi, M. & Sachdev, S. (2016). Stroke: a global response is needed. *Bulletin of the World Health Organization*. [Online]. Available at: http://dx.doi.org/10.2471/BLT.16.181636 [Accessed on 03 November 2019].

Julayanont, P., Phillips, N., Chertkow, H. & Nasreddine, Z.S. (2012). The Montreal Cognitive Assessment (MoCA): Concept and clinical review. *Cognitive Screening Instruments: A Practical Approach.* Springer-Verlag, pp. 111-152.

Kamwesiga, J.T., von Koch, L., Kottorp, A. & Guidetti, S. (2016). Cultural adaptation and validation of the Stroke Impact Scale 3.0 version in Uganda: A small-scale study. SAGE Open Medicine.4(1):1-10.

Katan, M. & Luft, A. (2018). Global burden of stroke. *Seminars in Neurology* 38(2):208-211. Available at: https://www.thieme-connect.com/products/ejournals/pdf/10.1055/s-0038-1649503.pdf [Accessed on 25 November 2019]

Katz, R.T., Rovai, G.P., Brait, C. & Rymer, Z. (1992) Objective Quantification of Spastic Hypertonia: Correlation With Clinical Findings. *Archives of Physical Medicine and Rehabilitation* 73(1): 339-347.

Kaya, T., Karatepe, A.G., Gunaydin, R., Koc, A. & Ercan, U.A. (2011). Inter-rater reliability of the Modified Ashworth Scale and modified Modified Ashworth Scale in assessing poststroke elbow flexor spasticity. International Journal of Rehabilitation Research. 34(1):59-64.

Kurtais, Y., Kucukdeveci, A., Elhan, A., Yilmaz, A., Kalli, T., Tur, B.S. & Tennant, A. (2009). Psychometric properties of the Rivermead Motor Assessment: Its utility in stroke. Journal of Rehabilitation Medicine. 41(1):1055-1061.

Li, F., Wu, Y. & Xiong, L. (2014). Reliability of a new scale for measurement of spasticity in stroke patients. Journal of Rehabilitation Medicine.46(1):746-753.

Lin, F.M. & Sabbahi, M. (1999). Correlation of spasticity with hyperactive stretch reflexes and motor dysfunction in hemiplegia. *Archives of Physical Medicine and Rehabilitation* 80(5):526-530.

Liston, R.A.L. & Brouwer, B.J. (1996). Reliability and validity of measures obtained from stroke patients using the Balance Master. *Archives of Physical Medicine and Rehabilitation 77*(5):425-530.

Maleka, M., Stewart, AS. & Hale, L. (2012) The experience of living with stroke in low urban and rural socioeconomic areas of South Africa. *South African Journal of Physiotherapy* [online]. 68(3): 25-29. Available at: https://sajp.co.za/index.php/sajp/article/view/21 [Accessed on 04 August 2018]

Mao, H.F., Hsueh, I.P., Tang, P.F., Sheu, C.F. & Hsieh, C.L. (2002). Analysis and comparison of the psychometric properties of three balance measures for stroke patients. *Stroke.* 33(1):1022-1027. Available at https://dx.doi10.1161/01.str.0000012516.63191.c5 [Accessed on 20 October 2019].

McDowell, I., Kristjansson, B., Hill, G.B. & Herbert, R. (1997). Community screening for dementia: The Mini Mental State Exam (MMSE) and modified Mini-Mental State Exam (3MS) compared. *Journal of Clinical Epidemiology.* 50(4):377-383.

Michalczak, M., Lewicki, M., Lagowska-Batyra, A., Bojarczuk, K. & Smolen, A. (2017) Evaluation of effectiveness of physiotherapy treatment on health improvement in patients after stroke. *Journal of Education, Health and Sport* [online]. 7(8): 1047-1061. Available at: http://dx.doi.org/10.5281/zenodo.1000455 [Accessed on 04 August 2018].

Min, J.H., Shin, Y.I., Joa, K.L., Ko, S.H., Shin, M.J., Chang, J.H. & Ko, H.Y. (2012). The Correlation between Modified Ashworth Scale and Biceps T-reflex and Inter-rater and Intra-rater Reliability of Biceps T-reflex. *Annals of Rehabilitation Medicine.* 36(4):538-543. Available at https://dx.doi.10.5535/arm.2012.36.4.538 [Accessed on 20 October 2019].

Mokkink, L.B., Prinsen, C.A.C., Patrick, D.L., Alonso, J., Bouter, L.M., de Wet, H.C.W. & Terwee, C.B. (2018). COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs). User manual version 1.0 dated February 2018. Available at: https://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf [Accessed on 08 August 2019].

Naghdi, S., Ansari, N.N., Raji, P., Shamili, A., Amini, M. Hasson, S. (2016). Cross-cultural validation of the Persian version of the Functional Independence Measure for patients with stroke. Disability and Rehabilitation. 38(3): 289-298.

Niama Natta DD , Batcho CS , Stoquart GG, Alagnidé E, Kpadonou T, Lejeune TM. (2019). Evaluation of manual ability in stroke patients in Benin: cultural adaptation and Rasch validation of the ABILHA ND-Stroke questionnaire. Eur J Phys Rehabil Med 2019;55:19-28.

Ostrofsky, C. & Seedat, J. (2016). The South African dysphagia screening tool (SADS): A screening tool for a developing context. South African Journal of Communication Disorders. 63(1):Art. #117, 9 pages.

Osundiya, O.C., Owolabi, M.O. & Hamzat, T.K. (2016). Sensitivity and Responsiveness of Ibadan Stroke-Specific Pain Scale. African Journal of Physiotherapy and Rehabilitation Science. 8(1&2):17-20.

Oveisgharan, S., Shirani, S., Ghorbani, A., Soltanzade, A., Baghaei, A., Hosseini, S. & Sarrafzadegan, N. (2006). Barthel Index in a Middle East Country: Translation, Validation and Reliability. Cerebrovascular Diseases. 22(1):350-354.

Parnes, P., Cameron, D., Christie, N., Cockburn, L., Hashemi, G. & Yoshida, K. (2009). Disability in low-income countries: Issues and implications. *Disability and Rehabilitation*. 31(14):1170-1180.

Pont, W., Groeneveld, I., Arwert, H., Meesters, J., Mishre, R.R., Vlieland, T.V. Goossens, P. & on behalf of the SCORE-study group. (2018). Caregiver burden after stroke: changes over time? *Disability and Rehabilitation. Available at:* https://doi.org/10.1080/09638288.2018.1499047 [Accessed on 25 November 2019].

Quinn, T.J., Langhorne, P. & Stott, D.J. (2011) Barthel Index for Stroke Trials. Development, Properties, and Application. *Stroke* [Online]. 42(1):1146-1151. Available at http://doi.org/10.1161/STROKEAHA.110.598540. [Accessed on 06 October 2019]

Rhoda, A., Smith, M., Putman, K., Mpofu, R., DeWeerdt, W. & DeWit, L. (2014) Motor and functional recovery after stroke: a comparison between rehabilitation settings in a developed versus a developing country. *BioMed Central Health Services Research* [Online]. 14(82):1-7. Available at: http://www.biomedcentral.com/1472-6963/14/82. [Accessed on 28 June 2018].

Sahathevan, R., Ali, K.M., Ellery, F., Mohamad, N.F., Hamdan, N., Ibrahim, N.M., Churilov, L. & Cumming, T.B. (2014). A Bahasa Malaysia version of the Montreal Cognitive Assessment: validation in stroke. International Psychogeriatrics. 26(5):781-786.

Sogbossi, E.S., Thonnard, J.L. & Batcho, C.S. (2014). Assessing Locomotion Ability in West African Stroke Patients: Validation of ABILOCO-Benin Scale. Archives of Physical Medicine and Rehabilitation.95(1): 1470-1476.

Soyuer, F. & Soyuer, A. (2004). Ischemic stroke: Motor impairment and disability with relation to age and lesion location. *The Internet Journal of Neurology*. 3(2):1-7. Available at doi:10.5580/9b4 [Accessed on 13 October 2019].

Stewart, A.L., Thrasher, A.D., Goldberg, J. & Shea, J.A. (2012). A framework for understanding modifications to measures for diverse populations. *Journal of Ageing and Health.* 24(6):992-1017.

Teasell, R. & Hussein, N. (2018). Clinical consequences of stroke. *Evidence based review of stroke rehabilitation* [Online]. Available at: http://www.ebrsr.com/sites/default/files/v18-SREBR-CH2-NET.pdf [Accessed on 25 November 2019].

Uniform Data System for Medical Rehabilitation. (2012). *The FIM Instrument: Its Background, Structure, and Usefulness.* New York, USA. Available at https://www.udsmr.org/Documents/The_FIM_Instrument_Background_Structure_and_Usefulness.pdf [Accessed on 06/10/2019].

Veras, M., Kairy, D., & Paquet, N. (2016). What Is Evidence-Based Physiotherapy? *Physiotherapy Canada [Online] 68*(2), 95–98. Available at https://doi:10.3138/ptc.68.2.GEE [Accessed on 23 November 2019].

Wade, D.T., Collen, F.M., Robb, G.F. & Warlow, C.P. (1987). Walking after stroke: Measurement and recovery over the first three months. *Scandinavian Journal of Rehabilitation Medicine*. 19(1): 25-30.

World Bank, The. (2018). *World Bank Country and Lending Groups*. [online] Available at: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [Accessed 23 June 2018].

World Bank Data Team (2019). New country classifications by income level: 2019-2020. Available at: http://blogs.worldbank.org/opendata/new-country-classifications-income-level-2019-2020 [Accessed on 23 October 2019].

World Health Organisation: Geneva (2002). Towards a common language for functioning, disability and health. Available at: https://www.who.int/classifications/icf/icfbeginnersguide.pdf Accessed on: 24 November 2019.

World Health Organization: Global Health Observatory. (c2019). Non-communicable diseases. Available from: https://www.who.int/en/news-room/fact-sheets/detail/noncommunicable-diseases Accessed on: 24 November 2019.

World Health Organization: Health Topics. (c2019). Stroke, Cerebrovascular accident. Available from: http://www.emro.who.int/health-topics/stroke-cerebrovascular-accident/index.html [Accessed 27 June 2018].

Wolf, S.L., Catlin, P.A., Gurucharri, K., Robertson, R. & Stephen, K. (1999). Establishing the reliability and validity of measurements of walking time using the Emroy Functional Ambulation Profile. *Physical Therapy.* 79(12): 1122-1133.

Yaliman, A., Kesiktas, N., Ozkaya, M., Eskiyurt, N., Erkan, O. & Yimaz, E. (2014). Evaluation of intrarater and interrater reliability of the Wisconsin Gait Scale with using the videotaped stroke patients in a Turkish sample. NeuroRehabilitation.34(1):253-258

Yan, L.L., Li, C., Chen, J., Marinda, J.J., Luo, R., Bettger, J., Zhu, Y., Feigin, V., O'Donnell, M., Zhao, D. & Wu, Y. (2016). Prevention, management, and rehabilitation of stroke in low- and middle-income countries. *eNeurologicalSCI.* 2:21-30. Available at: https://doi: 10.1016/j.ensci.2016.02.011 [Accessed on 28 November 2019].

Appendices

Appendix A: The World Bank Classification

Low Income	Lower middle income		Upper middle income	
Afganistan	Angola	Myanmar	Albania	Marshall Island
Benin	Armenia	Nicaragua	Algeria	Mauritius
Burkina Faso	Bangladesh	Nigeria	American Samoa	Mexico
Burundi	Bhutan	Pakistan	Argentina	Montenegro
Central African Republic	Bolivia	Papua New Guinea	Azerbaijan	Namibia
Chad	Cabo Verde	Philippines	Belarus	Nauru
Comoros	Cambodia	Sao Tome and Principe	Bosnia and Herzegovina	Panama
Congo, Dem. Rep	Cameroon	Solomon Island	Botswana	Paraguay
Eritrea	Congo, Rep.	Sri Lanka	Brazil	Peru
Ethiopia	Cote d'Ivoire	Sudan	Bulgaria	Romania
Gambia, The	Djibouti	Swaziland	China	Russian Federation
Guinea	Egypt, Arab Rep.	Syrian Arab Republic	Colombia	Samoa
Guinea-Bissau	El Salvador	Tajikistan	Costa Rica	Serbia
Haiti	Georgia	Timor-Leste	Croatia	South Africa
Korea, Dem. People's Rep	Ghana	Tunisia	Cuba	St. Lucia
Liberia	Guatemala	Ukraine	Dominica	St. Vincent and the Grenadines
Madagascar	Honduras	Uzbekistan	Dominican Republic	Suriname
Malawi	India	Vanuatu	Equatorial Guinea	Thailand
Mali	Indonesia	Vietnam	Ecuador	Tonga
Mozambique	Jordan	West Bank and Gaza	Fiji	Turkey
Nepal	Kenya	Yemen, Rep.	Gabon	Turkmenistan
Niger	Kiribati	Zambia	Grenada	Tuvalu
Rwanda	Kosovo		Iran, Islamic Rep.	Venezuela, RB
Senegal	Kyrgyz Republic		Iraq	
Sierra Leone	Lao PDR		Jamaica	
Somalia	Lesotho		Kazakhstan	
South Sudan	Mauritania		Lebanon	
Tanzania	Micronesia, Fed. Sts.		Libya	
Togo	Moldova		Macedonia, FYR	
Uganda	Mongolia		Malaysia	
Zimbabwe	Morocco		Maldives	

Adapted from: World Bank, The. (2018) World Bank Country and Lending Groups. [online] Available at: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [Accessed 23 June 2018].

Appendix B: Ethics approval

UNIVERSITEIT
STELLENBOSCH
UNIVERSITY

Approval Notice

New Application

11/09/2019

Project ID :10703

HREC Reference No: X19/08/032

Project Title: Outcome measures validated for use in the rehabilitation of stroke patients in low and middle income countries: A systematic review protocol.

Dear Ms Nabila Schoonraad,

The **New Application** received on 28/08/2019 10:46 was reviewed by members of **Health Research Ethics Committee** via **expedited** review procedures on 11/09/2019 and **was approved.**

Please note the following information about your approved research protocol:

Approval date: 11 September 2019

Expiry date: 10 September 2020

Please remember to use your Project ID 10703 and Ethics Reference Number X19/08/032 on any documents or correspondence with the HREC concerning your research protocol.

Please note that the HREC has the prerogative and authority to ask further questions, seek additional information, require further modifications, or monitor the conduct of your research and the consent process.

After Ethical Review

Translation of the informed consent document(s) to the language(s) applicable to your study participants should now be submitted to the HREC.

Please note you can submit your progress report through the online ethics application process, available at: Links Application Form Direct Link and the application should be submitted to the HREC before the year has expired. Please see *Forms and Instructions* on our HREC website (www.sun.ac.za/healthresearchethics) for guidance on how to submit a progress report.

The HREC will then consider the continuation of the project for a further year (if necessary). Annually a number of projects may be selected randomly for an external audit.

Provincial and City of Cape Town Approval

Please note that for research at a primary or secondary healthcare facility, permission must still be obtained from the relevant authorities (Western Cape Department of Health and/or City Health) to conduct the research as stated in the protocol. Please consult the Western Cape Government website for access to the online Health Research Approval Process, see: https://www.westerncape.gov.za/general-publication/health-research-approval-process. Research that will be conducted at any tertiary academic institution requires approval from the relevant hospital manager. Ethics approval is required BEFORE approval can be obtained from these health authorities.

We wish you the best as you conduct your research.

For standard HREC forms and instructions, please visit: Forms and Instructions on our HREC website https://applyethics.sun.ac.za/ProjectView/Index/10703

If you have any questions or need further assistance, please contact the HREC office at 021 938 9677.

Yours sincerely,

Mrs. Melody Shana

Coordinator

HREC1

National Health Research Ethics Council (NHREC) Registration Number:

Appendix C: Database search strings

Africa Journal Online

Type of search: Advanced Search

Within journals: All journals

Filters: None

Search string:

Search for:	
1.	Stroke AND assessment
2.	Stroke AND measure
3.	Stroke AND test
4.	Stroke AND physiotherapy
5.	Stroke AND validity
6.	Stroke AND valid*
7.	Stroke AND reliability
8.	Stroke AND “outcome measure”

AOSIS Publishing

Type of search: Simple search

Filters: None

Search string:

Search for:	
1.	Stroke AND assessment
2.	Stroke AND measure
3.	Stroke AND test
4.	Stroke AND physiotherapy
5.	Stroke AND validity
6.	Stroke AND valid*
7.	Stroke AND reliability
8.	Stroke AND “outcome measure”

BioMed Central

Types of search: Simple search: “Search all BCM articles”

Filters: None

Search string: (Stroke OR "cerebrovascular accident") AND (“outcome assessment” OR “outcome measure”) AND (validity OR reliability OR psychometric) AND (rehabilitation OR physiotherapy OR "physical therapy")

Cochrane Library

Type of search: Advanced search

Filters: None

Search string:

Search Manager:	
#1	MeSH descriptor: [Stroke] explode all trees
#2	MeSH descriptor: [Rehabilitation] explode all trees
#3	MeSH descriptor: [Outcome Assessment (Health Care)] explode all trees
#4	Validity OR reliability OR psychometric
#5	#1 AND #2 AND #3 AND #4

EBSCO Host: Africa-Wide Information & CINAHL

Type of search: Advanced search

Options selected: Suggest Subject Terms

Search string: Stroke AND Rehabilitation AND (“Outcome measure” OR “Outcome Assessment”) AND (Reliability OR Validity)

PEDro/Physiotherapy Choices

Type of search: Advanced search

Search field: Abstract & Title

Search string:

Search for:	
1.	Stroke AND assessment AND validity
2.	Stroke AND measure AND validity
3.	Stroke AND assessment AND reliability
4.	Stroke AND measure AND reliability
5.	Stroke AND assessment AND psychometric
6.	Stroke AND measure AND psychometric

ProQuest

Type of search: Advanced search

Filters: Source type: Scholarly Journals

Language: English

Location: Low, Lower middle, and upper middle income countries according to world bank classification

Search String: Stroke AND ("outcome measure" OR "outcome assessment") AND ("stroke rehabilitation" OR "neurological rehabilitation" OR rehabilitation) AND (validity OR reliability OR psychometric)

PubMed: MedLine

Type of search: Advanced search

Filters: None

Search string: "Stroke"[Mesh] AND ("Rehabilitation"[Mesh] OR "Stroke Rehabilitation"[Mesh] OR "Neurological Rehabilitation"[Mesh]) AND "Outcome Assessment (Health Care)"[Mesh] AND (validity OR reliability OR psychometric)

SABINET African Journals

Type of search:	Advanced search
Search string:	(Stroke OR “cerebrovascular accident”) AND ("outcome measure" OR "outcome assessment") AND ("stroke rehabilitation" OR "neurological rehabilitation" OR rehabilitation) AND (validity OR reliability OR psychometric)

Science Direct

Type of search:	Advanced search
Search Field:	Title, abstract or keywords
Filter:	Article type: Research articles only
Search string:	(Stroke OR "cerebrovascular accident") AND ("outcome assessment" OR "outcome measure") AND (validity OR reliability OR psychometric) AND (rehabilitation OR physiotherapy OR "physical therapy")

SCOPUS

Type of search:	Document search
Search selection:	All fields
Filter:	Location: Low, lower middle, and upper middle income countries according to world bank classification
Search string:	Stroke AND (“Rehabilitation” OR “Stroke Rehabilitation” OR “Neurological Rehabilitation”) AND “Outcome Assessment (Health Care)” AND (Validity OR Reliability OR Psychometric)

Web of Science

Type of search: Advanced search

Search string:

Search for:	
#1	TS=(Stroke OR “cerebrovascular accident”)
#2	TS=(“outcome assessment)
#3	TS=(“outcome measure”)
#4	#3 OR #4
#5	TS=(validity OR reliability OR psychometric)
#6	#5 AND #4 AND #1

Appendix D: Critical Appraisal Tool by Brink & Louw (2011).

	Item criteria	Result	Scoring criteria
1	Was the participant sample described in detail?	Yes	Description of participants such as age, gender, diagnosis.
		No	Not described in detail.
		N/A	Not human participants.
2	Was the competence of index test rater described?	Yes	Description of raters such qualification, competence with outcome measure.
		No	Not described in detail.
3	Was an explanation of the reference standard provided?	Yes	Description of reference standard in terms of content, appropriate and accuracy.
		No	Not described in detail.
4	Was inter-rater blinding ensured?	Yes	Raters unaware of the results of other raters.
		No	Not described in detail.
		N/A	Not relevant.
5	Was intra-rater blinding ensured?	Yes	Raters unaware of their own results.
		No	Not described in detail.
		N/A	Not relevant.
6	Was randomisation of the test order ensured?	Yes	Test order not systematic.
		No	Not described in detail.
		N/A	Not relevant.
7	Was the interval between the index and reference standard tests sufficient?	Yes	Interval sufficient to avoid change in variable.
		No	Allows change in variable.
		N/A	Not relevant.
8	Was the stability of the variable ensured between test periods?	Yes	Interval sufficient to avoid change in variable.
		No	Allows change in variable/Not relevant.
9	Was the reference standard included in the index test?	Yes	Test developed from another.
		No	Two different tests.
10	Was the index test procedure described in detail?	Yes	Testing process described in detail.
		No	Not described in detail.

11	Was the reference standard test procedure described in detail?	Yes	Testing process described in detail.
		No	Not described in detail.
12	Was any participants unaccounted for (withdrawals)?	Yes	All participants accounted for.
		No	Withdrawals not explained.
		N/A	Not human participants.
13	Was the psychometric test applicable?	Yes	Best methods used for analysis.
		No	Inappropriate test used.

Adapted from: Brink, Y. & Louw, QA. (2011) Clinical instruments: reliability and validity critical appraisal. *Journal of Evaluation in Clinical Practice*, [online] 18(6): 1126-1132. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2753.2011.01707.x [Accessed 22 June 2018].

Appendix E: Data Extraction Form

Demographic data:

Authors	Title	Year	Country	Income Classification	Study Aims	Purpose	Sample				Study Setting
							Size	Age	Gender	Diagnosis	

Measurement tool data:

Outcome measure	Short title	Outcome	ICF Domain	Reference standard	Translation language	Cross-cultural adaptation process

Psychometric properties:

Reliability		
Internal Consistency	Test re-test; Interrater; Intrarater	Measurement error: Test re-test; Interrater; Intrarater

Validity		
Content validity	**Construct validity**	**Criterion Validity**

Validity			
Face validity	**Structural validity**	**Hypothesis testing**	**Concurrent validity; Predictive validity**

Responsiveness	Interpretability	CAT score	Citation/Reference