DEVELOPMENT OF A CLASSIFICATION SCALE FOR ARM USE FOLLOWING STROKE

by

Lisa Ann Simpson

BASc., Queen’s University, 2000
MSc.OT., Queen’s University, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
in
THE FACULTY OF GRADUATE STUDIES
(Rehabilitation Sciences)
THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

March, 2013

© Lisa Ann Simpson, 2013
Abstract

Introduction and Purpose

Understanding upper extremity functional recovery following stroke requires knowledge of how individuals use their upper extremity in the home and community settings. Currently, there are no measures that provide a rich description about the types of activities and manner in which they are performed by the affected upper extremity. This study aimed to develop and validate a classification scale that captures UE use following stroke.

Objectives

1. To conduct a systematic review of existing research on the ability of UE function and UE use measures following stroke to capture important, observed and detectable change.
2. To develop a new classification tool, Rating of Everyday Arm-use in the Community and Home (REACH) scale, that captures UE use and to assess the interrater reliability and validity of the new tool.

Methods

Design: This study used multiple methods to achieve the stated objectives. A systematic review methodology that reviewed 68 articles covering 14 measures was employed. Development of the REACH scale utilized focus groups to acquire client, clinician, and caregiver perspectives. Evaluation of interrater reliability involved a test-retest design with different raters. Hypothesis testing was used to assess the validity of the REACH scale.

Participants: Focus groups included clinicians (n=13), participants with stroke (n=16), and caregivers (n=4). Individuals with stroke living in the community participated in the evaluation of the REACH scale (n=96).
Measures: Responsiveness indices were extracted and calculated from the included articles for the systematic review. The following measures were used to evaluate the REACH scale: Motor Activity Log, activity counts captured by accelerometers, Action Research Arm Test, Stroke Impact Scale – hand subscale, Chedoke-McMaster Stroke Assessment (Arm Recovery, Hand Recovery, and Shoulder pain scales).

Results
The REACH scale, informed by literature and user consultation, takes 5 minutes to administer. The intraclass correlation coefficient and weighted kappa for interrater reliability was 0.96 and 0.91 respectively. Strong relationships were observed between the REACH scores and external measures of UE use, UE function and UE impairment (ρ=0.91-0.94).

Conclusions
The REACH scale provides an accessible and efficient tool for measuring UE use following stroke.
Preface

In collaboration with my committee (Dr. Janice Eng, Dr. William Miller and Dr. Catherine Backman), I developed the research program, collected and analysed the data, and was primary author on all chapters of this thesis.


80% Contribution: In collaboration with Dr. Eng, I conceptualized the study and developed the research design. I was the primary person responsible for conducting the literature search, analysing the data and writing the manuscript. Dr. Eng provided input on data analysis, interpretation of results and editing of the manuscript.

A version of Chapter 3 will be submitted for publication. Simpson LA, Eng JJ, Miller WC, Backman CL. Rating of Everyday Arm-use in the Community and Home (REACH) Scale: development and validation of a classification scale for arm-use following stroke.

80% contribution: In collaboration with Dr. Eng, I conceptualized the research study and design. Dr. Miller and Dr. Backman also provided input and feedback regarding the study design. I was the primary person responsible for data collection, analysis and drafting of the manuscript. Dr. Eng provided input on data analysis and interpretation of results and was the primary editor of the manuscript. All authors provided feedback on the presentation of the results and all authors reviewed and edited the manuscript.
The project received approval from the Behavioural Research Ethics Board of the University of British Columbia (certificate number: H10-03180).
Table of Contents

Abstract ................................................................................................................................. ii
Preface ................................................................................................................................. iv
Table of Contents ............................................................................................................... vi
List of Tables ....................................................................................................................... x
List of Figures ...................................................................................................................... xi
Acknowledgements ........................................................................................................... xii
1. Introduction ....................................................................................................................... 1
  1.1 Epidemiology of stroke ............................................................................................... 1
  1.2 Epidemiology and impact of upper extremity dysfunction ....................................... 1
  1.3 UE function and the ICF ............................................................................................ 2
  1.4 Discrepancy between UE functional capacity and UE functional performance .......... 2
  1.5 The importance of UE use ......................................................................................... 4
  1.6 Measurement of UE Use ............................................................................................. 4
  1.7 Classification scales in clinical research ...................................................................... 6
  1.8 Development and evaluation of an UE use classification scale ............................. 8
    1.8.1 Scale development ............................................................................................... 8
    1.8.2 Scale evaluation ................................................................................................. 10
      1.8.2.1 Validity .......................................................................................................... 10
      1.8.2.2 Reliability ..................................................................................................... 12
    1.8.3 Factors that impact UE use following stroke .................................................... 14
  1.9 Rationale, objectives and hypothesis of thesis ......................................................... 18
    1.9.1 Chapter 2: a systematic review of the responsiveness of UE functional measures following stroke ............................................................... 18
    1.9.2 Chapter 3: development and validation of a classification scale to capture UE use following stroke ............................................................. 19
2. Systematic review of the responsiveness of UE functional measures following stroke: a focus on important, detectable and observed change ........................................ 21
  2.1 Introduction .................................................................................................................... 21
  2.2 Methods ......................................................................................................................... 22
    2.2.1 Literature search ............................................................................................... 22
2.2.2 Inclusion criteria ........................................................................................................ 23
2.2.3 Data extraction and organization ............................................................................. 24
2.3 Results .......................................................................................................................... 25
  2.3.1 Observed change ...................................................................................................... 25
  2.3.2 Important change .................................................................................................... 27
  2.3.3 Detectable change .................................................................................................... 29
2.4 Discussion ...................................................................................................................... 30
  2.4.1 Observed change ..................................................................................................... 31
  2.4.2 Important change .................................................................................................... 32
  2.4.3 Detectable change .................................................................................................... 33
2.5 Conclusion ..................................................................................................................... 35
Bridging statement .............................................................................................................. 36
3. Rating of Everyday Arm-use in the Community and Home (REACH) scale: development and evaluation of a classification scale for arm-use following stroke ............................................................... 37
  3.1 Introduction .................................................................................................................. 37
3.2 Methods ........................................................................................................................ 38
  3.2.1 Development of the REACH scale ......................................................................... 38
    3.2.1.1 Goals for REACH development ....................................................................... 38
    3.2.1.2 Overall development process ......................................................................... 39
    3.2.1.3 Recruitment ..................................................................................................... 39
    3.2.1.4 Procedure ....................................................................................................... 40
    3.2.1.5 Data analysis .................................................................................................. 41
    3.2.1.6 Formation of the REACH scale ...................................................................... 42
  3.2.2 Evaluation of the REACH Scale ............................................................................. 43
    3.2.2.1 Design .............................................................................................................. 43
    3.2.2.2 Participants .................................................................................................... 43
    3.2.2.3 Measures ....................................................................................................... 44
    3.2.2.4 Procedures .................................................................................................... 47
    3.2.2.4 Data analysis .................................................................................................. 48
3.3 Results .......................................................................................................................... 51
Appendix J: Stroke Impact Scale-hand ................................................................. 145
Appendix K: Chedoke-McMaster Stroke Assessment (arm, hand and shoulder scales) . 146
List of Tables

Table 1.1 Relevant questions for assessment of clinical sensibility .................................70
Table 1.2 Five aspects of construct validity and their associated sources of evidence ...71
Table 2.1 Number of articles for included measures by category of change ...................72
Table 2.2 Minimally clinically important difference (MCID) values calculated by
distribution and anchor-based methods ...........................................................................73
Table 3.1 Focus group participant characteristics .............................................................74
Table 3.2 REACH evaluation participant characteristics ..................................................75
Table 3.3 Distribution of REACH scale scores ..................................................................76
Table 3.4 Shapiro Wilk test of normality for the REACH scores .....................................77
Table 3.5 Interrater reliability and measurement error values.........................................78
Table 3.6 Distribution of clinical measures used in validity analyses .........................79
Table 3.7 Correlation coefficients for validity analysis ....................................................80
List of Figures

Figure 1.1 Sigmoidal relationship between UE use and UE function .................................. 81
Figure 1.2 COSMIN taxonomy of measurement properties .................................................. 82
Figure 1.3 Conceptualization of validity for thesis ................................................................. 83
Figure 2.1 Graphic representation of observed, important and detectable change .......... 84
Figure 2.2 Flow diagram of process to select final list of outcome measures ............... 85
Figure 2.3 Effect sizes by measure calculated at <3 months and ≥ 3 months post stroke .......................................................... 86
Figure 2.4 Comparison of observed change captured by ‘lab-based’ versus ‘patient-perceived functional measures .......................................................... 87
Figure 2.5 Comparison of MDC₉₅ values relative to the sample means and scale maximums ............................................................................. 88
Figure 3.1 Rating of Everyday Arm-use in the Community and Home (REACH) scale .................................................................................. 89
Figure 3.2 Plot of rater agreement ......................................................................................... 93
Figure 3.3 Scatterplot of REACH scale scores with external measures of UE use, UE function and UE impairment .......................................................... 94
Acknowledgements

I could not have completed this research without the advice, assistance, and support of many important people. The first, and biggest, thank you goes to my supervisor and mentor Janice Eng. Janice has supported me in immeasurable ways from our first meeting through the final moments I spent preparing this manuscript. I am grateful for her time, expertise, and encouragement. In our relatively short time together I have begun to internalize her guidance and advice. When I am writing or working on my research, I often feel as though Janice is in the room with me and I am certain my work is better for it. I also want to thank my committee members. Both Bill Miller and Catherine Backman have provided enthusiastic and thoughtful feedback along the way. It was a great pleasure to work with them.

I am also extremely grateful to Chihya Hung for her assistance with many stages of the research – especially her help with participant recruitment and assessments. I truly enjoyed working with Chihya on this project and aspire to adopt her cheerful and efficient approach to all things in my future research. Thank you also to Amanda, Christie and Amy for their help with assessments and to Amanda for her help with data preparation.

Of course I could not have completed this project without the support of my husband Andrew. I owe you weeks of dog walks and dinners.

Over the past year I have also benefited, both academically and personally, from my colleagues: Krista, Paula, Ada, Jeremy, Brodie, Dom, Debbie, Bita and Ed. Thank you for your support and friendship.

I am sincerely grateful to the many individuals who participated in this study. I learned a great deal working with them. Without willing and able volunteers, research such
as this would not be possible. Thanks to all of these people for generously contributing their time.

Finally, thank you to my mom and dad for always supporting me in all my endeavors. This one is for you Dad.
1. Introduction

1.1 Epidemiology of stroke

Approximately 50,000 strokes occur annually in Canada (Heart and Stroke Foundation of Canada 2012) and currently over 300,000 Canadians are living with the effects of stroke (Public Agency of Canada 2009). The annual cost of stroke to the Canadian economy is over 2.8 billion dollars with the majority of these costs driven by hospitalization and rehabilitation care (Mittman et al. 2012). Individuals with greater disability are associated with a two-fold increase in costs due to longer lengths of hospital and rehabilitation stays (Mittman et al. 2012).

1.2 Epidemiology and impact of upper extremity dysfunction

Upper extremity (UE) hemiparesis is particularly common with up to 70% of individuals experiencing this impairment following a stroke (Nakayama et al.1994). Moreover, it has also been observed that despite receiving rehabilitation, over 65% of individuals experience difficulty incorporating their affected upper limb into daily activities (Broeks et al.1999). Facilitating functional changes in the UE is a major goal of rehabilitation following a stroke. Increased UE function is associated with shorter length of stay (Persson et al. 2012), increased ability to perform functional activities (Morris et al. 2013; Sveen 1999), increased participation in life activities (Harris and Eng 2007; Sveen 1999) and increased quality of life (Morris et al. 2013; Mayo et al. 1999). In addition, findings from qualitative studies with stroke survivors have highlighted the important role of using the UE in life activities; acting as both a motivating factor to continue training (Timmermans et al. 2009) and a contributor to the perception of recovery (Barker and Brauer 2005).
1.3 UE function and the ICF

The International Classification of Functioning, Disability and Health (ICF) is a useful tool for understanding the different levels of functioning that an injury or disease can impact (WHO 2001). UE function lies within the activity level of the ICF, which is defined as the completion of a task or action by an individual. The ICF outlines two important qualifiers of the activity domain: 1) activity capacity and 2) activity performance. Activity capacity refers to what an individual is capable of doing under controlled conditions whereas activity performance refers to what an individual actually does in his or her own environment.

1.4 Discrepancy between UE functional capacity and UE functional performance

Despite demonstrating sufficient UE activity capacity as captured by lab-based measures, studies have shown that individuals with stroke still demonstrate decreased UE activity performance, as captured by measures that assess arm use in one’s own environment (Han et al. 2013; Rand and Eng 2012; Dromerick et al. 2006; Van der Lee et al. 2004; Sterr et al. 2002). The relationship between UE activity capacity (herein referred to as UE function) and UE activity performance (herein referred to as UE use) is complex. Correlations between measures assessing function and measures assessing use range from 0.4-0.82 (Chen et al. 2012; Michielson et al. 2009; Harris et al. 2007; Lang et al. 2007; Dromerick et al. 2006; Uswatte et al. 2006b; Van der Lee et al. 2004; Van der Pas et al. 2001). More recent research exploring the relationship between use and function suggest it is best modeled as a non-linear relationship (Hidaka et al. 2012; Schweighofer et al. 2009; Han et al. 2008). Using computational models and data derived from a large clinical trial that examined the effect of Constraint-Induced Therapy, Hidaka et al. (2012) found that a
sigmoidal function better modelled the relationship between UE function and UE use than a linear function (Figure 1.1). This same research group demonstrated preliminary support for the existence of a functional threshold by fitting sigmoidal computer models to data obtained on the large Constrain-Induced Therapy trial (Schweighofer et al. 2009). The model generated by Schweighofer et al. (2009) suggests that surpassing a particular functional threshold after therapy leads to a long term increase in spontaneous use. If the functional threshold isn’t achieved, spontaneous use decreases. Only 68% of the people above the obtained functional threshold showed an increase in use however. The authors acknowledge that many more factors likely contribute to long term spontaneous use of the UE and thus explain the low predictive capability of their calculated functional threshold.

Finally, a discussion concerning the relationship between UE function and UE use would not be complete without mention of the theory of non-use. This theory was used to explain how a behavioural program (called Constraint-Induced therapy), which consisted of restraining the non-affected UE combined with intensive task-practice of the affected UE, resulted in an increase in affected UE use (Taub 2006). Thus, the basic theory explains that observed non-use of the affected UE following a stroke is primarily a result of a learned behaviour. Essentially, individuals with stroke learn to compensate for their initial motor impairments by using their non-affected arm and continue to do so despite improvements in motor function. A recent article by Hidaka et al. (2012), which used computational models to explain non-use, suggested the following factors were important in explaining the learned non-use phenomenon: 1) decreased function of the affected UE; 2) relative increase in function of the unaffected UE (especially if the dominant side was affected) and 3) decreased confidence in using the affected UE for a particular task.
In summary, the findings from studies that examine the relationship between UE function and UE use suggest that UE use is a complex and independent construct. The observed discrepancy between arm function and arm use is reasonable as measures of arm function capture an individual’s best effort at one time whereas measures of arm use capture what a person is able to do across a period of time while taking into consideration the individual’s wants, needs and environment demands. Thus understanding function in daily life requires the understanding of more than just capacity.

1.5 The importance of UE use

Decreased arm use can lead to increased pain (Eng et al., 2008) weakness (Eng et al., 2008) decreased bone density (Eng et al., 2008) and decreased quality of life (Williams et al. 1999). Thus, research investigating UE functional recovery following stroke must consider outcomes that reflect how the affected upper extremity is being used in one’s own environment. Moreover, understanding whether people with stroke eventually go on to use their affected UE once they return to the community is important given the considerable amount of resources devoted to facilitating UE function.

1.6 Measurement of UE Use

Recent reviews examining outcome measures that capture UE function following stroke have identified a scarcity of tools that capture affected arm use (Connell and Tyson 2012; Lemmens et al. 2012; Baker et al. 2011; Chen and Weinstein 2009; Ashford et al. 2008). Outcome measures are any physical instrument, self-report or observational tool that captures the construct under study. Currently accelerometry and the Motor Activity Log (MAL) (Uswatte et al. 2006b; Van der Lee et al. 2004; Uswatte et al. 2005) are the only two
measures that capture how much the affected UE is being used in one’s own environment. The MAL is a self-report measure that was developed as an evaluative tool to capture change in arm use following Constraint-Induced Therapy (Van der Lee et al. 2004). Respondents are asked to rate how much and how well they use their affected UE during 30 activities of daily living (e.g. brushing teeth, holding a book, using towel to dry self). Each activity is rated on two separate Likert scales: 1) Amount of Use scale ranging from 0 (I didn’t use my weaker arm) to 5 (I used my weaker arm as often as before the stroke) and 2) Quality of Movement scale ranging from 0 (My weaker arm was not used at all for that activity) to 5 (The ability to use my weaker arm for that activity was as good as before the stroke). Correlations between the amount of use scale and the quality of movement scale are high with coefficients ranging from 0.8-0.95 (Van der Lee 2004; Uswatte et al. 2005; Uswatte et al. 2006b). The final MAL scores are an average rating across all of the activities. Thus, summary scores range from 0 to 5 with higher scores representing higher use. A short form MAL also exists which consists of 14 activities. The recommended administration time for the full version is one hour (Taub et al. 2011). Reported test-retest reliability ICC values range from 0.79-0.82 and limits of agreement values range from 12-17% of the scale range (-0.61 to 0.85). A two week interval in between testing sessions was used to obtain these values (Uswatte et al. 2006b; Van der Lee et al. 2004). Validity has been supported through correlations between the MAL and accelerometry (r=0.47-0.70) and MAL and measures of UE function (r=0.63-0.72) (Uswatte et al. 2006b; Uswatte et al. 2005; Van der Lee et al. 2004). Although, the clinical feasibility is rated as excellent with respect to its cost, portability and training required (Connell and Tyson 2011), clinical interpretation of the summary score is not straightforward. Indeed, as it was developed to be an evaluative tool, the summary score provides little qualitative
information about how the arm is being used. For example, a summary score of 3 out of 5 does not provide information about the nature of the use (e.g. grasp, reach, stabilization).

Accelerometry is an objective measure that captures movement of the arm by measuring its acceleration and converting it into an activity count (Rand and Eng 2010). Test-retest reliability (using an interval time period of 2 weeks) has been supported with ICC values ranging from 0.82-0.94 (Uswatte et al. 2006a). Evidence for validity was provided from the following sources: obtaining expected differences in accelerometer values between known groups (Lang et al. 2007; Uswatte et al. 2006a); correlations between accelerometry and the MAL (r=0.41-0.51) (Uswatte et al. 2006a); and correlations between accelerometry and measures of impairment (r=0.54-0.75) (Gebruers et al. 2008). Accelerometers capture both functional and non-functional movement and thus activity counts or duration of arm use cannot provide information about the types of activities and manner in which they are performed.

In summary, the literature has provided evidence to support adequate validity and reliability of the two measures that currently capture UE use. However, neither the MAL nor accelerometry is able to provide valuable descriptive information about how the affected UE is used in clients’ daily lives.

1.7 Classification scales in clinical research

Classification scales that are quick and easy to administer have made positive contributions in various clinical research fields. The modified Rankin Scale, Hoehn Yahr Scale and the Manual Ability Classification Scale are three examples of widely-used classification scales to assess ability in stroke, Parkinson’s, and cerebral palsy populations, respectively.
The modified Rankin Scale was developed in the 1980s and is used as a descriptive and evaluative tool to assess disability following stroke (New and Buchbinder 2006). The scale consists of 7 levels which range from “no symptoms” to “death”. It is the most frequently used outcome measure among stroke trials published in high impact international journals (Quinn et al. 2009) and has been recommended for use in acute stroke clinical trials (Lees et al. 2012). A major strength of this scale is its ability to capture meaningful change. Lai and Duncan (2001) found that transition from one level to another was associated with a meaningful change in disability. Other cited strengths of the modified Rankin Scale are its short administration time, the depiction of patient outcomes in real-world settings, its association with other meaningful outcomes such as quality of life and economic measures and its discriminatory power (Lees et al. 2012).

The Hoehn Yahr scale is a classification scale originally developed in the 1960s as a staging scale to describe clinical function in Parkinson’s disease (Hoehn and Yahr, 1967). The Hoehn Yahr scale is the most widely used scale to describe the severity of Parkinson’s disease worldwide (Goetz et al. 2004). Indeed, the Google Scholar citation quote for the seminal Hoehn and Yahr (1967) article surpasses 6000 citations. Cited strengths of this scale include its easy and quick administration, wide use, stages that provide predictive value and practicality in both research and clinical settings (Goetz et al. 2004).

Finally, the Manual Ability Classification Scale is a more recent classification scale developed in 2006 (Eliasson et al. 2006). It is a descriptive tool that classifies UE ability in children with Cerebral Palsy. Although developed recently it is already recognized as a useful tool for classifying motor abilities among individuals with a heterogeneous and difficult to diagnose condition (Morris 2007).
A classification scale that provides valuable descriptions of and distinctions in everyday UE use has the potential to capture meaningful UE recovery and make an important contribution to the field of stroke rehabilitation.

1.8 Development and evaluation of an UE use classification scale

1.8.1 Scale development

Development of a classification scale that captures UE use following stroke can be guided by a field known as clinimetrics. Clinimetrics was originally developed to apply standards and procedures to the measurement of “inherently clinical phenomenon” (Feinstein 1987). Examples of clinical phenomenon include functional status, quality of life and pain. Essentially a clinical phenomenon is a construct that cannot always be captured through laboratory procedures or represented by a ratio numerical value (Feinstein 1982). According to clinimetrics, common attributes of scales used to assess clinical phenomena include: a) capture a complex clinical phenomenon (e.g. quality of life, pain); b) have an ordinal or nominal scale; c) capture a multidimensional construct (e.g. participation); d) select appropriate ratings via observation, interview or self-report; e) perform basic clinical functions such as describing status, documenting changes, making predictions or inform clinical guidelines; and f) consider clinical sensibility (Feinstein 1987). The author defines clinical sensibility as a mixture of common sense, knowledge of pathophysiology and clinical reality. Clinimetrics is a suitable framework to guide the development of an UE use classification scale as the attributes outlined above describe many of the desired properties of the new scale.

Feinstein (1987) outlines guidelines for the development of multidimensional clinical scales that meet the requirements of clinical sensibility. The guidelines involve the following
steps: 1) selection of potential component dimensions/variables; 2) retention of component dimensions/variables; 3) combination of component dimensions/variables and 4) formation of the output scale. Selection of candidate variables (Step 1) can be accomplished through conceptualizing the construct or clinical phenomenon and through review of background information. Construct definition and background information can come from the literature, clinical expert opinions and client experiences (Baker et al. 2011; Streiner and Norman 2008; Feinstein 1987). Retention of component variables (Step 2) requires a decision about the relative importance of the potential variables and the goals of the scale (ie. easy to use versus comprehensive representation of construct). The sources of information used to generate the potential variables can also be used to guide the decisions about their relative importance (Feinstein 1987). Feinstein (1987) asserts that successful completion of Steps 3 and 4, and ultimately adherence to attributes of clinical sensibility, depends on the scope, discrimination and coherence of the final scale. Other important areas that contribute to a scale’s clinical sensibility relate to the scale’s purpose, overt format, face validity, content validity and ease of use. Table 1.1 provides relevant questions within these areas to help scale users evaluate scale sensibility.

Application of the clinimetric process to the development of a classification scale that captures meaningful levels of UE use following stroke leads to the formulation of the following goals:

1) Determination of the most pertinent factors and levels that meaningfully describe and distinguish UE use. The following sources can help identify and narrow down the most pertinent factors: clinicians, client groups and the literature.
2) Creation of UE use categories that are mutually exclusive, cover a realistic range of possible use categories, meaningfully discriminate between levels of use, are adequately described and are clinically reasonable.

3) Formulation of a scale administration process that is clear and quick.

1.8.2 Scale evaluation

There are multiple resources in the literature that can guide the evaluation of a health measurement scale. Although these resources are not often complimentary, examination of multiple resources can assist in formulation of a comprehensive evaluation framework. For instance, standards that have been created in order to assess the quality of measurement studies can help guide the evaluation process. The COSMIN checklist is an example of guidelines created to assess the quality of measurement studies (Mokkink et al. 2010b). The checklist was developed through international expert consensus and provides a taxonomy of the major measurement properties to address in the evaluation of a health related measure. Figure 1.2 outlines the COSMIN taxonomy of measurement properties. In this thesis, the following properties will be evaluated: validity (construct) and reliability (interrater reliability and measurement error).

1.8.2.1 Validity

Validity is defined as a “judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores” (Messick 1989). It is a continual process that involves the accumulation of evidence and is not an all or nothing property of the test (Mokkink et al. 2010a; Messick 1989). Validity was traditionally broken down into the following major types of validity:
content validity, criterion validity and construct validity (Streiner and Norman 2008).

Contemporary views of validity however have shifted based on the paradigm introduced by Messick (1989). Messick (1989) describes validity as a unified concept that is centred on hypothesis testing. Hypotheses are generated from expectations of how scores should behave based on the underlying construct the test is purported to measure. In essence, all forms of validity can be tested based on a construct validity framework.

Messick (1995) proposes 5 different aspects of construct validity and their associated sources of evidence that can guide the validation process. The aspects are: content, substantive, structural, external and consequential. The five aspects are outlined in Table 1.2. The COSMIN framework for assessment of validity also describes different forms of evidence that guide the validation process (Figure 1.2). Although the terminology differs, many of the COSMIN definitions adhere to a unified concept of validity. To avoid confusion, this thesis will conceptualize and evaluate validity based on Messick’s unified definition. Figure 1.3 represents the conceptualization of validity adopted in this thesis.

When formulating hypotheses for evaluation of construct validity, the COSMIN guidelines recommend that if using correlation to investigate a hypothesized relationship between the new scale and another scale, the magnitude and direction of the correlation should be specified (Mokkink et al. 2010a). The authors also recommend specifying the direction and magnitude for hypotheses based on expected mean differences between groups. Finally, standards for hypothesis testing outlined in the COSMIN checklist also states that the p-value is not as relevant to the judgement about the quality of evidence in support of the hypothesis. The p-value is a reflection of the sample size and not the expected relationships outlined in the hypotheses (Mokkink et al. 2010a). Challenges arise when the construct on
which hypotheses are based is not well-defined or the theoretical underpinnings of the construct are still largely unknown (Messick 1989). Therefore, the ongoing process of validation often involves the combination of both exploratory and confirmatory hypotheses (Messick 1989).

1.8.2.2 Reliability

Reliability reflects the consistency of a tool and the extent to which it is free from error (Portney and Watkins 2009). Measurement error can be introduced from various sources. For instance, error can be introduced if an assessment is administered by different raters and their interpretations of the responses do not agree. Exploration of the error attributed to two or more raters administering the scale to the same group of subjects is referred to as interrater reliability and will be reported in Chapter 3 (Portney and Watkins 2009). A common design for assessing interrater reliability is having different raters simultaneously and independently test a subject. Simultaneous testing is not always possible or practical in situations when the interaction between the tester and the subject is essential (Portney and Watkins). The interaction between the tester and the client was thought to be an essential part of the newly developed classification scale outlined in Chapter 3. Testing trials using different raters therefore needed to occur at separate times. Consideration of an appropriate time frame in between subsequent testing trials was therefore an important part of the interrater reliability design. When selecting an appropriate time frame in between testing trials, the time period should be short enough so that the underlying construct has not changed and long enough to prevent recall bias of either the rater or subject (Streiner and Norman 2008). Participants who are greater than 6 months post stroke are considered stable.
and thus change in upper extremity use is not expected within time frames from 1-3 weeks (Van der Lee 2004).

By their nature, most classification scales are ordinal scales. Recommendations regarding the most appropriate reliability coefficient to use for ordinal/categorical scales are conflicting (Connell and Tyson 2012; Mokkink et al. 2010a; Streiner and Norman 2008). Measurements of agreement such as kappa are often recommended for categorical/ordinal scales (Portney and Watkins 2009). A kappa statistic is a measurement of agreement that considers the number of observed agreements and compares this to the number of agreements obtained through chance. Thus the final kappa value represents the proportion of agreements that can be attributed to reliable measurement (Portney and Watkins 2009). The kappa statistic does not distinguish between disagreements however. For instance, a researcher may want to differentiate between measurements that disagree by one category versus three. The weighted kappa was thus introduced to account for different types of disagreements (Sim and Wright 2005). This statistic allows researchers to assign weights to specific types of disagreements. The most common types of weights are linear and quadratic (Sim and Wright 2005). The COSMIN guidelines recommend researchers use the weighted kappa coefficient for ordinal scales and report the type of weighting system used (Mokkink et al. 2012a). Reporting of the weighting system is important as the weights can have an effect on the kappa statistic. For instance, kappa values increase with the number of categories when quadratic weights are used (Brenner and Kliebsch 1996). This is especially the case when the number of categories ranges from 2-5. The magnitude of the kappa coefficient is less affected by the number of categories when linear weights are used (Brenner and Kliebsch 1996). Thus comparisons of weighted kappas are difficult when
comparing scales with a small (2-5) and differing number of categories. It has been observed that weighted kappa is equivalent to the intraclass correlation coefficient (ICC) when using the quadratic weighting system (Brenner and Kliebsch 1996; Streiner and Norman 2008; Portney and Watkins 2009). This observation adds to the complexity in the choice of reliability for ordinal scales. Streiner and Norman (2008) argue that the kappa coefficient should only be adopted for nominal scales with two categories. Otherwise, they recommend the use of the ICC. As the ICC is the most widely used reliability coefficient and the newly developed scale is on the borderline of having a small number of categories (6), both the weighted kappa and the ICC will be used to capture interrater reliability in Chapter 3.

1.8.3 Factors that impact UE use following stroke

Reviewing the literature for factors associated with a construct under investigation provides important background information for scale development and aids in formulating relevant hypotheses for the validation process. Despite the complex nature of affected UE use following stroke, there are few studies that have investigated the factors associated with use. Studies that examine predictors of UE recovery rely on UE functional capacity as the end point recovery measure (Chen and Weinstein 2009). Initial motor capability and initial function were the best predictors of UE recovery according to recent systematic reviews of predictors of UE recovery following stroke (Coupar et al. 2012; Chen and Weinstein 2009). Motor capability can be captured through measures of muscle strength, active range of motion or composite motor impairment measures such as the Fugl-Meyer Assessment or the Chedoke McMaster Stroke Assessment (Chen and Weinstein 2009). Other frequently studied clinical predictors of UE function include spasticity, decreased sensation, pain, global stroke severity, side of stroke and cognition (Cougar et al. 2012; Chen and Weinstein 2009). The
The most frequently studied demographic predictors were age, sex and time post stroke (Cougar et al. 2012).

The relationship between UE use and many of these clinical and demographic factors have also been investigated in the limited number of studies that involve UE use as an outcome. Appendix A summarizes the relationships between arm use and the most frequently examined clinical and demographic factors. The factors are organized according to the ICF into impairments, function and personal factors.

Over half of the studies (11/19) examining factors associated with UE use were small with sample sizes less than 50 participants (Han et al. 2013; Hidaka et al. 2012; Thrane et al. 2011; Van der Pas et al. 2011; Michielson et al. 2009; Rinehart et al. 2009; Gebruers et al. 2008; Reiterer et al. 2008; Lang et al. 2007; Dromerick et al. 2006; Pang et al. 2006). Nine studies used the MAL to capture UE use while nine studies used accelerometry. One study used a new measure called the Bilateral Arm Reaching Test (BART) to assess UE use and non-use.

UE motor capability was the most frequently examined factor among studies that examined the relationship between UE impairments and UE use (N=8). One study used a measure of active range of motion (Lang et al. 2007), three studies used a measure of strength (Harris et al. 2007; Lang et al. 2007); and five studies used a composite measure of UE motor impairment (Gebruers et al. 2008; Lin et al. 2009d; Michielson et al. 2009; Pang et al. 2006; Thrane et al. 2011). All measures of motor capability were significantly associated with UE use in all eight studies. The correlations between measures of active range of motion/strength ranged from 0.37-0.52. Reiterer et al. (2008) found that strength was significantly correlated to UE use at one day and one week post stroke and not significantly
correlated to use at 3 and 6 months post stroke. This is in contrast to Harris and Eng (2007) who found that strength explained 78% of the variance in UE use among individuals 5 years post stroke. Correlations between measures of composite UE impairment and UE use ranged from 0.54-0.75 in cross-sectional studies (Thrane et al. 2011; Michielson et al. 2009; Gebruers et al. 2008). Increased motor capability as captured by measures of composite impairment was also related to greater UE use following exercise or CIMT treatments (Lin et al. 2009d; Pang et al. 2006).

Fewer studies assessed the relationship between UE use and impairments such as UE pain, decreased sensation, spasticity and stroke severity and these results were more heterogeneous. Spasticity was significantly correlated to UE use ($\rho$=-0.71, $p<0.01$) in the one study that assessed both spasticity and UE use; however was not a significant factor within a multiple linear regression model (Harris and Eng 2007). Sensation was significantly correlated to UE use ($\rho$=-0.34, $p<0.01$) in one out of two studies that examined this factor; however was not uniquely related to UE use in a multiple linear regression model (Harris and Eng 2007). UE pain was significantly correlated to UE use in one study which involved individuals with acute stroke (Lang et al. 2007) however was not related to UE use in the other study which involved individuals 5 years post stroke (Harris and Eng, 2006). Finally, a relationship between stroke severity and UE use was not consistently found among the three studies in which it was assessed (Lin et al. 2009d; Gebruers et al. 2008; Reiterer et al. 2008). Only one study found a significant correlation between stroke severity and UE use ($\rho$=-0.59 to -0.75, $p<0.001$) (Gebruers et al. 2008).

UE function was significantly associated with UE use in all eleven studies that examined the relationship between these two constructs. Studies that examined the
correlations between functional measures and use found coefficients ranging from 0.4 to 0.82 (Chen et al. 2012; Michielson et al. 2009; Harris et al. 2007; Lang et al. 2007; Dromerick et al. 2006; Uswatte et al. 2006b; Van der Lee et al. 2004; Van der Pas et al. 2001). One study used regression models to determine the relationship between function and use and found that function explained 30-66% of the variance in use (Michielson et al. 2009). Interestingly, following the input of several variables into a multiple linear regression model to explain improvement in UE use following CIMT treatment, only post-test function and a measure of strength remained in the model. These factors explained 12.3% of the variance in long-term improvement of use (Schweighofer et al. 2009). This further supports the theory that an increase in spontaneous use is supported if individuals surpass a functional threshold following therapy. Function and strength provide a small percentage of this increase in use however providing further evidence of the complexity of UE use following stroke.

Finally, hand dominance or concordance of hand dominance and side of stroke was the most common personal factor examined in studies examining factors related to UE use (N=7 studies). Side of hand dominance was not significantly associated with UE use when using the MAL as the measure of use (Schweighofer et al. 2009; Dromerick et al. 2006; Fritz et al. 2006) however it was related to patterns of UE use among the studies that used accelerometry (Haaland et al. 2012; Rinehart et al. 2009; Uswatte et al. 2006a). For instance, patterns of unilateral use and bilateral use differed depending on whether the dominant UE or non-dominant UE was affected (Han et al. 2013; Haaland et al. 2012; Rinehart et al. 2009). In addition correlations between self report measures of UE use (i.e. the MAL) and objective measures of UE (ie. accelerometry) were greater among the dominant UE affected individuals than the non-dominant UE affected individuals (r=0.56-0.59 vs r=0.28-0.34).
Review of the influences on UE use revealed three significant factors: UE motor capability, UE function and side of hand dominance in relation to side of stroke. These findings influenced the development of the classification scale and design of its evaluation. For instance, separate scales were developed depending on whether the dominant side or non-dominant side was affected. In addition, different types of tasks that the affected UE performed were used to differentiate between levels in the developed classification scale. The complexity of the tasks corresponded to the levels of the scale such that higher levels on the use scale corresponded to the performance of more complex tasks. Thus one would expect motor capability and function to correlate with the scores on the classification scale. These hypotheses were included in the validation component of Chapter 3.

1.9 Rationale, objectives and hypothesis of thesis

The overall purpose of this thesis is to develop and validate a classification scale that captures UE use following stroke. Two studies were conducted in order to accomplish this overall purpose and are outlined in Chapters 2 and 3. The following section outlines the rationale, objectives and hypotheses of these chapters.

1.9.1 Chapter 2: a systematic review of the responsiveness of UE functional measures following stroke

Rationale: One of the major goals of developing a classification scale that captures UE use following stroke is to formulate categories that cover a realistic range of possible UE use levels and discriminates between levels of UE use that are meaningful to the individual with stroke. Achievement of this goal requires an understanding of what meaningful UE use looks like. Meaningful or important change is considered one type of change that is captured by an
instrument. The type of change captured by an instrument is said to influence a scale’s responsiveness (Beaton et al. 2001). Beaton et al. (2001) created a responsiveness taxonomy that includes three major types of change: important, observed and detectable. Using Beaton’s taxonomy of responsiveness the different types of change captured by measures that assess UE function or UE use will be reviewed.

**Objectives:** To conduct a systematic review of the ability of UE function and UE use measures following stroke to capture important, observed and detectable change.

**1.9.2 Chapter 3: development and validation of a classification scale to capture UE use following stroke**

**Rationale:** There are currently no outcome measures that capture valuable information about how the UE is being used in a person’s own environment following stroke. A new tool is thus needed that assesses this independent and complex aspect of UE recovery. Classification scales can provide meaningful information to guide or evaluate therapy and are quick and easy to use. In addition, evidence supporting adequate reliability and validity is necessary in order to confidently use a newly developed tool.

**Objectives:** 1) To develop a new classification scale that captures UE use following stroke in one’s own environment. 2) To assess the interrater reliability and validity of the new tool.

**Hypothesis:**

The newly developed classification scale will:

1) Have good inter-rater reliability (ICC≥0.80; weighted kappa≥0.60)
2) Be correlated to existing measures of UE use, UE function and UE impairment with correlation coefficients greater than 0.5.
2. Systematic review of the responsiveness of UE functional measures following stroke: a focus on important, detectable and observed changea

2.1 Introduction

Functional recovery following stroke is complex with wide variation in natural recovery and response to treatment across individuals. Optimizing or augmenting changes in recovery is core to the rehabilitation process following a stroke. Hence, it is essential that outcome measures are able to detect change as it occurs; a property known as responsiveness (Mokkink et al. 2010; Beaton et al. 2001).

Detecting change over time or from an intervention is one of the most critical requisites of an outcome measure; it is necessary information for selecting the best instrument for practice or research and for determining sample size for clinical studies. Furthermore, there is growing recognition that clinical studies should report treatment effects that are considered meaningful to their study population (Kain and MacLaren 2007). Thus in addition to traditional reporting of statistical significance tests and effect sizes studies should include methods for determining meaningful or important change.

Beaton et al. (2001) argue that responsiveness is a context-specific characteristic that is influenced by factors such as the specific sample, treatment, and the type of change captured by an instrument. The authors outline three major types of change in their responsiveness taxonomy. They are: 1) observed change 2) important change and 3) detectable change. Figure 2.1 depicts the inter-relationship between the three types of change and the common metrics used to quantify them. Observed change is the amount of change

---

a A version of this chapter has been published. The final, definitive version of this paper has been published in Neurorehabilitation and Neural Repair, 27, March/April/2013 by Sage Publications Ltd/Sage Publications, Inc., All rights reserved. © Simpson LA, Eng JJ. Functional recovery following stroke: capturing changes in upper-extremity function. 2013;27:240-250. The article can be found at: http://nnr.sagepub.com/content/27/3/240.
observed in a population in which change is expected to occur (i.e. after a treatment of known efficacy or a specific period within the natural recovery pattern). Traditional methods have captured this type of change with an effect size (Beaton et al. 2001). Important change is the observed change estimated to be meaningful and is often quantified as a minimal clinically important difference value (MCID). For example, important change may reflect the value that patients, clinicians or society places on the recovery (or partial recovery) of a task, like the ability to bring a spoon to the mouth. Finally, detectable change takes into consideration the measurement error associated with a tool and is often quantified as a minimal detectable change score (MDC) or limit of agreement (LOA). Calculation of the MDC value associated with a measure varies depending on the confidence level selected. The most common confidence levels selected are 95% and 90% and are denoted by the subscripts MDC\textsubscript{95} and MDC\textsubscript{90}. Confidence that true functional change has been observed in a clinical study is increased when the observed change is equal to or surpasses a measure’s detectable change values (Beaton et al. 2001).

The purpose of this paper is to synthesize and critically review the research evidence that captures the responsiveness of UE functional measures following stroke as defined by three types of change (observed, important, detectable).

2.2 Methods

2.2.1 Literature search

This review targeted articles that capture three types of change (observed, important, detectable) using UE functional outcome measures following a stroke. Articles were identified using a systematic search of electronic databases (MEDLINE, EMBASE, CINAHL, PsycINFO, Cochrane CENTRAL) from database inception through March 2012.
The following keywords were used: stroke, cerebrovascular accident or hemiplegia or hemiparesis combined with upper extremity, function or activities of daily living and responsiveness, reliability, psychometrics, “minimal clinically important change or MCID”, “standard error of measurement or SEM”, “minimal detectable change or MDC”, “standardized response mean or SRM”, effect size or outcome measurement. All terms were mapped onto subject headings. Articles were limited to the English language, human subjects, and adults. A hand search of reference lists from reviews and the grey literature (e.g. Stroke Engine Assess) was also conducted to ensure a thorough search (Hopewell et al. 2007).

2.2.2 Inclusion criteria

To be included in this review, articles had to meet the following inclusion criteria: 1) provided a responsiveness index or sufficient information to calculate responsiveness index (e.g. minimally detectable change or effect size); 2) utilized a sample of individuals with stroke; 3) used an outcome measure that assessed affected UE ability and included ≥50% functional activities (to ensure scale reflected International Classification of Functioning, Disability and Health activity domain) and 4) published in a peer-reviewed journal.

Exclusion criteria consisted of: 1) conference proceedings or abstracts; 2) pre-post studies or randomized control trials (RCTs) which calculated a responsiveness index without utilizing the control group information to minimize bias for inflation of the effect sizes (Morris and Deshon 2002); 3) studies with a sample size less than ten; 4) articles that utilized measures that captured UE function as a single dimension (e.g. Box and Blocks test, peg test). Single dimensions were not considered representative of the many dimensions involved in UE
function. RCTs were excluded if they: 1) utilized <2 UE functional outcome measures or 2) found no significant effects for UE function.

2.2.3 Data extraction and organization

Responsiveness indices were extracted or calculated from the included articles and subsequently organized into the three categories of change outlined by Beaton et al. (2001): 1) observed change, 2) important change and 3) detectable change (Figure 2.1). Observed change was further subdivided into 1) change over natural recovery (categorized into <3 months post injury and ≥ 3 months post injury) and 2) change in response to an effective treatment. Effect sizes were calculated based on the change score divided by the baseline standard deviation and minimally detectable change values ($MDC_{90}$, $MDC_{95}$) were calculated based on the test-retest coefficient (ICC) and the baseline standard deviation (Stratford 2004). In addition, $MDC_{95}$%, which are independent of measurement units, were calculated to compare minimal detectable change values across measures (Flansbjer et al. 2005). We present the $MDC_{95}$% using the two methods commonly cited in the rehabilitation literature: 1) $MDC_{95}$/maximum score for the scale (e.g. 57 points for ARAT) and 2) $MDC_{95}$/baseline mean of the sample.

RCTs that utilized at least two different functional outcome measures were used to estimate observed change in response to a treatment of known efficacy. This allowed for comparison of observed change across different measures within the same study by controlling for variation in treatment and sample characteristics (Angst 2011). To ensure we were capturing observed change in response to an effective treatment, only RCTs that found a significant effect for at least one of the UE functional measures were included. Scatterplots of effect sizes were generated for a visual representation of one outcome measures’ relative
ability to capture change compared to another measure in response to the same treatment with the same sample. When only median and range scores were provided, mean and standard deviation values were estimated using the method suggested by Hozo et al. (2005) in order to calculate an effect size.

2.3 Results

The search strategy yielded 1770 titles of which 68 met the inclusion criteria (Figure 2.2). References for the included articles can be found in Appendix B. The articles provided responsiveness data for 14 functional outcome measures. Table 2.1 displays the number of articles that provided responsiveness estimates for each measure. The Action Research Arm Test, Motor Activity Log, Wolf Motor Function test and Stroke Impact Scale were the four most frequently used measures among the included articles.

2.3.1 Observed change

Observed change over natural recovery

Effect sizes that captured change over time of participants in standard care were extracted or calculated from 25 studies. Fourteen studies followed participants up to 3 months post stroke and ten studies followed participants up to 6 months post stroke. As only one study followed participants past 6 months post stroke, we compared the effect sizes for participants less than 3 months versus those greater or equal to 3 months post injury. The majority of studies had observation periods from 2 to 5 months.

References

Effect sizes calculated at a baseline of 1-3 months post injury were larger and showed greater variance than effect sizes calculated at a baseline ≥ 3 months post injury (Figure 2.3). This finding was evident when the duration of follow-up was similar between the two groups. Two studies (Pandyan et al. 2003; Duncan et al. 1999) considered the effect of stroke severity on observed changes over recovery; effect sizes were 2.0-2.6 times larger for individuals with less severe impairments at 1-2 months post-stroke at study baseline. Three studies (Rabadi and Rabadi 2006; Higgins et al. 2005; Desrosiers et al. 2003) calculated effect sizes using two methods: population effect size (based on the change score divided by baseline standard deviation) and standardized response mean (based on the change score divided by the change score standard deviation) for the same measure. In all three studies, the effect sizes calculated as a standardized response mean were larger than when calculated as a population effect size (Appendix C).

**Observed change in response to a treatment of known efficacy**

A total of 28 RCTs utilized more than one UE functional outcome measure and obtained a significant effect for at least one of these measures. These 28 RCTs were used to examine observed change of different measures in response to a treatment of known efficacy within the same study. Constraint-Induced Therapy was the most frequently studied intervention among the included RCTs (50%). Also, the majority of these studies utilized a lab-based performance measure (e.g. Action Research Arm Test, Wolf Motor Test), in addition to a

---

measure that captured perceived function in one’s own environment (e.g. Motor Activity Log, Stroke Impact Scale) (n=23). The most common outcome measures used together were the Motor Activity Log with the Action Research Arm Test (used in 7 studies)\textsuperscript{d} or the Motor Activity Log with the Wolf Motor Test (used in 10 studies).\textsuperscript{e}

The majority of effect sizes were close to a 1:1 relationship between the lab-based versus patient-perceived functional measures within the same study (Figure 2.4). Effect sizes from studies which investigated Constraint Induced Therapy (Hsu et al. 2010; Myint et al. 2008; Sun et al. 2010; Taub et al. 2006; Wittenberg et al. 2003) and one study that examined the effect of mirror therapy for individuals with Complex Regional Pain Syndrome (Cacchio et al. 2009) did not demonstrate this 1:1 relationship. Effect sizes for the perceived effect (Motor Activity Log) were 1.6-6.2 times larger than the functional changes (measured by Action Research Arm Test or Wolf Motor Test) in these studies.

2.3.2 Important change

Five studies established important change (MCID) values for six functional outcome measures (Table 2.2).\textsuperscript{f} All but one study calculated MCID values from Constraint-Induced Therapy trials (Lin et al. 2010; Lin et al. 2009; Lang et al. 2008; Fritz et al. 2007). The other study investigated change following robotic therapy (Wang et al. 2011). Four studies (Wang et al. 2011; Lin et al. 2010; Lin et al. 2009; Fritz et al. 2007) utilized individuals with chronic stroke and one study sample consisted of individuals with acute stroke (Lang et al. 2008).

\textsuperscript{d} Shindo et al. 2011; Hsu et al. 2010; Sun et al. 2010; Harris et al. 2009; Myint et al. 2008; Page et al. 2005; van der Lee et al. 1999

\textsuperscript{e} Khan et al. 2011; Wu et al. 2011; Tariah et al. 2010; Cacchio et al. 2009; Dahl et al. 2008; Gauthier et al. 2008; Kowalczewski et al. 2007; Pang et al. 2006; Taub et al. 2006; Wittenberg et al. 2003

\textsuperscript{f} Wang et al. 2011; Lin et al. 2010; Lin et al. 2009b; Lang et al. 2008; Fritz et al. 2007
Three studies (Wang et al. 2011; Lin et al. 2010, Lin et al. 2009) used a combination of anchor and distribution based techniques whereas two studies (Lang et al. 2008; Fritz et al. 2007) used only anchor based methods to define important change. Anchor-based methods compare the change scores on the measure of interest to a comparative measure or ‘anchor’ of important change (Lydick and Epstein 1993). The following anchors were used in the five studies: 1) a predetermined level on a global rating scale in which participants were asked to rate their perception of functional change (Lang et al. 2008); 2) a predetermined level of recovery on the Stroke Impact Scale global recovery item (Wang et al. 2011; Lin et al. 2010; Fritz et al. 2007) and 3) a change score of 6-10 points on the upper extremity portion of the Fugl-Meyer assessment (Lin et al. 2009b). Distribution based methods determine important change based on the statistical distribution of the results (Lydick and Epstein 1993). Three studies used 0.2 times the standard deviation of the sample (i.e. effect size of 0.2) to determine MCID values (Wang et al. 2011; Lin et al. 2010; Lin et al. 2009b).

Important change values displayed large variation with values spanning from 1.1%-30% of the tests’ maximum scores. The largest values were observed in the study that utilized a sample of individuals in the acute stage post stroke (Lang et al. 2008). In addition, MCID values calculated using statistically-derived distribution based methods were substantially smaller (15-88% less) than the values determined by anchor based methods (Table 2.2). For example, the MCID for the Stroke Impact Scale-hand subscales ranged from a value of 5.8 using an effect size method (distribution method) to 17.8 using perceived amount of recovery on the global recovery question (participant-perceived anchor-based method) for the same sample of individuals (Lin et al. 2010).
Three studies were not able to establish MCID values due to non-significant relationships between the global recovery/rating scales and functional changes using the Motor Activity Log (Fritz et al. 2007; van der Lee et al. 2004), Wolf Motor Test (time component) (Lang et al. 2008; Fritz et al. 2007), and accelerometry (Lang et al. 2008) (Table 2.2).

2.3.3 Detectable change

Values needed to surpass measurement error, which are considered to represent true functional change (SEM, MDC$_{90}$, MDC$_{95}$ and LOA values), were extracted or calculated from 16 studies for 9 measures.$^8$ All studies used a test-retest methodology in which time frames in between assessments ranged from 1 day to 2 weeks. Nine studies utilized individuals with subacute stroke$^h$ and seven studies used samples with chronic stroke.$^i$ Nine studies utilized a subsample of individuals from Constraint Induced Therapy/forced use trials$^j$ of which four were from the EXCITE trial (Fritz et al. 2009; Uswatte et al. 2006a; Wolf et al. 2005).

Minimal detectable change values at the 90% and 95% confidence levels ranged from 1.0% and 1.2% of the maximum score for the Arm Motor Ability Test-time subscale to 21.9% and 25.9% of the maximum score for the Stroke Impact Scale-hand respectively. Relative to their sample means, minimal detectable change values at the 90% and 95% confidence levels ranged from 1.0% and 1.2% of the maximum score for the Arm Motor Ability Test-time subscale to 21.9% and 25.9% of the maximum score for the Stroke Impact Scale-hand respectively.


$^i$ Lin et al. 2010; Lin et al. 2009b; Lin et al. 2009c; Carod-Artal et al. 2008; Uswatte et al. 2005; van der Lee et al. 2004; van der Lee et al. 2001


29
confidence levels ranged from 11.5% and 13.7% for the Wolf Motor Test-functional ability subscale to 72.5% and 86.7% for the Motor Activity Log-amount of use subscale respectively. Detailed SEM, MDC$_{90}$ and MDC$_{95}$ estimates and their respective relative percent values (ie. relative to the sample mean or relative to the scale maximum score) can be found in Appendix D. Among the measures with multiple estimates, the values needed to surpass measurement error for tools that capture patient-perceived function (e.g. Motor Activity Log, Stroke Impact Scale) were larger than lab-based performance measures (e.g. Wolf Motor Test, Chedoke Arm and Hand Activity Inventory) (Figure 2.5). This observation was present whether the minimal detectable change values were considered relative to the scale maximums or the sample means.

Also, the values needed to surpass measurement error for patient-perceived performance measures were larger or on par with important change values. For instance, minimal detectable change and important change values for the Stroke Impact Scale were: 17.1-21.9 (MDC$_{90}$) and 20.4-25.9 (MDC$_{95}$) versus 5.8 (distribution-based MCID) or 17.8 (anchor-based MCID). Minimal detectable change and important change values for the Motor Activity Log were: 0.56-1.06 (MDC$_{90}$) and 0.67-1.27 (MDC$_{95}$) versus 1.0-1.1 (anchor-based MCID).

2.4 Discussion

Examining the ability of UE functional measures to capture observed, important and detectable change revealed several novel findings related to the ability of these outcome measures to capture functional recovery.
2.4.1 Observed change

For equal duration of follow up, the effect sizes due to natural recovery calculated at 1-3 months post stroke were substantially larger than those calculated at 3 months or later post stroke. The observed differences in effect sizes between these time phases likely reflect the higher degree of neuroplasticity that has been documented early after stroke (Biernaskie et al. 2004). The effect sizes obtained in the RCTs using a population at >6 months post stroke ranged from 0.05 to 4.28 demonstrating that individuals are still capable of change at later time periods post stroke when receiving treatment. Of importance, our collective data demonstrate that rehabilitation treatments can affect patient perceptions of functional change as effectively as lab-based functional measures. In fact, Constraint Induced Therapy is an exemplary treatment model where patient perceptions of change (Motor Activity Log) were 1.6-6.2 times larger than effect sizes obtained with lab-based functional performance measures (Action Research Arm Test or Wolf Motor Test). A likely explanation is that Constraint Induced Therapy was specifically designed to overcome learned non-use, thereby targeting functional change in one’s own environment. Additionally, the higher relative effect sizes observed in the Motor Activity Log following Constrain Induced Therapy could also reflect patient bias. This bias could be present due to the large investment of time and effort required from the Constraint Induced Therapy program. Finally, it should be noted that our observations concerning the relative ability of treatments to affect lab-based versus patient-perceived function may be limited to studies able to capture statistically significant effects.

Real time accelerometry monitoring of UE activity in the home and community is a promising technology to objectively capture function in one’s own environment without the
necessity of self-report (Dobkin and Dorsch 2011). However, only one study provided an estimate of true or important change for accelerometry measures. More studies are needed to better understand the usefulness of real time activity monitors for capturing the effectiveness of UE rehabilitation interventions.

The influence of the method used to calculate effect sizes (ie. methods based on baseline standard deviation versus standard deviation of the change scores) was also revealed. Methods based on the change score standard deviation (ie. standardized response mean) produced estimates that were up to 1.2-1.9 times larger than those calculated based on the baseline standard deviation (ie. population effect size). The method of effect size calculation is an important contextual factor that must be considered when designing and interpreting research. This has particular importance for interpreting treatment effects and when performing sample size calculations. The observed influence of calculation method also highlights the difficulty of using an effect size in isolation when making a judgement about a measure’s general responsiveness (Mokkink et al. 2010a).

2.4.2 Important change

A key finding was that important change values obtained through anchor-based methods (eg, based on patient opinion or comparative measures) were higher than those for the distribution methods (eg, statistical estimates) among the studies that used both approaches (Wang et al. 2011; Lin et al. 2010; Lin et al. 2009b). In fact, the MCID of one measure (Stroke Impact Scale) tripled in magnitude from the distribution to anchor-based approach using the same subjects (Lin et al. 2010). Our results indicate that distribution methods result in smaller MCID and researchers may be tempted to interpret their findings in light of this MCID, especially if the intervention has small effects. However, statistically-
driven distribution methods have been criticized for their lack of meaning to participants (Revicki et al. 2008; Beaton et al. 2001). The MCID from anchor-based approaches are larger, and may provide a challenge in finding therapies that can achieve this effect. Some have questioned whether it is realistic to achieve MCIDs derived from patient-perceived global rating scales (anchor-based) because many factors (e.g., recall bias, baseline characteristics, expectations of treatment and question format) can affect patients’ perception of change (Lang et al. 2008; Guyatt et al. 2002). It has been suggested that a combination of anchor–based methods from patient and clinical perspectives be used to determine a MCID value and distribution-based methods should only supplement this information (Revicki et al. 2008).

2.4.3 Detectable change

This synthesis highlighted important differences between lab-based and participant-perceived functional measures. Measures that capture perceived function in one’s own environment (e.g. Stroke Impact Scale, Motor Activity Log) required larger values to surpass their measurement error than lab-based performance measures (e.g. Wolf Motor Test, Action Research Arm Test). Although incorporating patient perspectives of functional change is an important component of capturing meaningful outcomes in neurorehabilitation research (Salter et al. 2005), researchers should be aware however of the larger sample size required to be able to capture true change using perceived function measures. Quality criteria guidelines recommend that the values needed to surpass measurement error calculated at a 95% confidence interval (ie. true change captured by MDC_{95}) should be less than the minimum values considered to be important (ie. MCID) (Terwee et al. 2007). In contrast, our study found that the MDC_{95} and MDC_{90} for patient-perceived functional measures were
similar or greater than the MCID. There remains debate in the literature however concerning the best estimate of minimal detectable change values for self-report measures. Some researchers argue that MDC_{90} and MDC_{95} produce overly conservative estimates and recommend setting the minimal detectable change to one SEM (Turner et al. 2010; Wywich et al. 1999).

One of the limitations of this paper is there is no one standard approach for conceptualizing responsiveness. We utilized the broad taxonomy described by Beaton et al. (2001). One international group (COSMIN) defined a narrower conceptualization of responsiveness which used a longitudinal validity approach such as correlating change with an external criterion or determining the area under the receiver operating characteristic (ROC) curve to distinguish between known groups (Mokkink et al. 2010). In addition, while there are established appraisal guidelines for intervention studies (e.g., PeDRO Score), consensus has yet to be reached with regard to assessing the rigour of studies that measure an outcome measure’s psychometric properties. Application of the Beaton taxonomy, which focuses on the nature of the change, provided a useful framework for understanding the state of the stroke literature in regards to the broader topic of measuring functional change. Effect sizes provided in this study can inform hypotheses for future responsiveness testing. Finally, ceiling and floor effects were also not considered in this review. Large ceiling and/or floor effects may indicate a subsection of the population for which measures are not as responsive and is another factor that can bias estimates of important change values (Stucki et al. 1996). For instance, floor and ceiling effects may indicate an underrepresentation of items at the lower and higher end of the scales. Capturing change at these ends may therefore be more difficult. Given the contextual nature of responsiveness, the dominance of Constraint
Induced Therapy among the included studies should be noted. Samples from Constraint Induced Therapy trials made up 80% of the important change articles, 56% of the detectable change articles and 50% of the RCTs used to examine observed change in response to an effective treatment. Thus, the literature informing the ability of measurement tools to capture functional change in the upper extremity post stroke may be biased to those individuals who are eligible for Constraint Induced Therapy trials.

2.5 Conclusion

In summary, this synthesis revealed important findings that have implications for the measurement and interpretation of upper extremity functional recovery following stroke. The magnitude of important change or change that surpasses measurement error varied substantially depending on the method of calculation used. Our findings suggest that rehabilitation treatments can affect patient perceptions of UE functional change as effectively as lab-based UE functional measures; however research studies may require higher sample sizes to account for the larger measurement error associated with patient-perceived functional measures. Future studies examining meaningful change in upper extremity function in varied subgroups of individuals (ie. at different levels of stroke severity and stage of recovery) are needed.
**Bridging statement**

The systematic review of the responsiveness of UE functional measures following stroke provides important background information for the development of an UE use classification scale. The study findings confirmed that there are presently only two outcome measures that capture UE use (MAL, Accelerometers). It also revealed that little information is known about what people consider to be meaningful change in UE use following stroke. Minimal Clinically Important Difference values were not determined for accelerometry and Minimal Clinically Important Difference values for the Motor Activity Log were calculated from only one study (Lang et al. 2008). Seeking opinions from individuals with stroke concerning their concepts of meaningful UE use is thus an essential first step in the development of a new UE use classification scale. In addition, the systematic review also found that measurement error associated with self-report UE functional measures was higher than that for performance based measures. This finding was taken into consideration when designing the reliability testing of the new classification scale.
3. Rating of Everyday Arm-use in the Community and Home (REACH) scale: development and evaluation of a classification scale for arm-use following stroke

3.1 Introduction

There is growing recognition that patient-oriented outcomes such as meaningful functional recovery should be used in the evaluation of stroke care (Reeves et al. 2010). Traditionally, upper extremity measures have focused on “capacity”, that is what an individual is capable of doing under controlled conditions (e.g., grip strength; Box and Block Test). Understanding functional recovery following stroke also requires knowledge of “performance”, how individuals actually use their affected upper extremities in their own environment. Therefore, outcomes that assess functional recovery should consider the activities the client undertakes on a regular basis in the home and community (Lang et al. 2012; Dobkin et al. 2011; Baker et al. 2011).

Currently, few outcome measures capture actual use of the affected UE in the community and home setting (Lemmens et al. 2012; Baker et al. 2011; Chen and Weinstein 2009). Accelerometry (Uswatte et al. 2006a) and the Motor Activity Log (MAL) (Uswatte et al. 2006b) are two measures that assess UE use. Studies utilizing these measures have shown that individuals with stroke often demonstrate little use of their paretic UE in their home and community, even if they have good capability to use their arm and hand (Han et al. 2013; Rand and Eng, 2012; Dromerick et al. 2006; Van der Lee et al. 2004; Sterr et al. 2002). However, both these tools have limitations that prevent their widespread adoption in clinical and research settings. For instance, wrist accelerometry, which objectively captures amount of UE use outside the clinical setting, can be expensive and the resulting activity counts or duration of arm use cannot provide information about the types of activities and manner in
which they are performed. The MAL is a 30-item self-report measure that was developed as an evaluative tool to capture change in arm use following Constraint-Induced Therapy (Van der Lee 2004). The scale is time consuming to administer (recommended time of one hour) and the summary score ranging from 0 to 5 does not provide any descriptive information about how the arm is being used.

We aimed to develop a quick and easy-to-administer classification scale that provides meaningful descriptions and distinctions of UE use outside the clinical setting. This scale could assist in setting realistic treatment goals, activity prescription and evaluation of patient-oriented outcomes. Indeed, existing classification scales have already served as valuable tools for capturing meaningful descriptions of and outcomes for individuals with neurological conditions (Quinn et al. 2009, Goetz et al. 2004). The purpose of this study was to 1) consult with clients, caregivers and clinicians on the development of a relevant UE use classification scale called the Rating of Everyday Arm-use in the Community and Home (REACH) and 2) quantify aspects of its validity and interrater reliability.

3.2 Methods

3.2.1 Development of the REACH scale

3.2.1.1 Goals for REACH development

The primary goal for development of the REACH scale was to create an ordinal level scale that would describe and distinguish between meaningful levels of affected UE use in the natural environments of individuals with stroke (ie. create mutually exclusive, clinically relevant and meaningful categories that cover the full range of use). The intent of the scale was to capture spontaneous UE use during daily activities, rather than the person’s capacity for UE performance. Finally, low administrator and respondent burden (ie. short
administration time, clearly worded and easy to score) was another desired property of the REACH scale. In summary, the goals for the REACH development process were to create a scale with attributes that adhere to principals of clinical sensibility as outlined in clinimetric principles (Feinstein 1987).

3.2.1.2 Overall development process
A multi-phase process was used to develop the number and content of the categories for the REACH classification scale. First, a combination of focus groups and interviews were conducted with clinicians, individuals with stroke and caregivers. Secondly, focus group findings were supplemented with information from the literature to generate an initial draft of the scale. Lastly, scale refinement occurred through feedback sessions with clinicians and pilot testing on individuals with stroke. This multi-phase process was used to achieve the goals stated above and ultimately promote the scale’s clinical sensibility. More specifically, this process was designed to promote the scale’s relevant content, adequate scope, coherence and ease of use. All procedures were approved by the University of British Columbia and Vancouver Coastal Health ethics boards.

3.2.1.3 Recruitment
Purposive sampling was used to recruit healthcare providers, individuals with stroke and their caregivers for the focus group sessions. Healthcare providers who worked in an outpatient or community setting with at least one year experience were invited to participate. Clinicians from the following disciplines were recruited: Occupational Therapy, Physical Therapy, Rehabilitation Medicine, Nursing and Recreational Therapy. Individuals with stroke with a range of UE functional impairment (as captured by active range of motion) and who met the following inclusion criteria were invited to participate: 1) ≥ 19 years old; 2) at
least 6 months post stroke and 3) live in the community and 4) have residual UE impairment due to the stroke (as captured by self-report). Caregivers of the individuals with stroke who met the above criteria were invited if they provided assistance with daily or instrumental daily activities.

3.2.1.4 Procedure
Separate sessions were conducted with 1) clinicians and 2) clients with stroke and their caregivers due to differences in the goals for these distinct groups. The clinician focus groups centred on generating the following information: 1) the important dimensions of UE use from the perspective of healthcare providers; 2) how the dimensions could be described in order to distinguish between levels of use and 3) the potential number of levels required to capture the full continuum of UE use. Group sessions with individuals with stroke and their caregivers focused on understanding meaningful UE use from client perspectives and more specifically, how UE use was described by clients and caregivers. Interviews were conducted for those individuals unable to attend group sessions or who had expressive aphasia which limited their communication in a group setting. Focus group and interview sessions were led by the author who utilized a discussion guide (Appendix E). An assistant moderator in group sessions took notes, aided in summarizing key points and assisted with clarifying meaning of participants’ statements. The moderator and assistant moderator debriefed following each group session and modified the discussion guide to account for emerging themes or difficulties. All interview and group sessions were audio recorded and transcribed. Participant characteristics were also collected. Informed written consent was obtained from all participants.
Clinicians who indicated an interest in providing feedback on the developed scale were sent the initial draft. Consultation meetings were set up to discuss the scale’s content, clarity, scaling and scoring procedures. Feedback from these consultations was used to make changes before pilot testing on a group of individuals with stroke. Preliminary testing of the scale was conducted on eight individuals with stroke who possessed a range of UE abilities and stroke characteristics. Former focus group participants were utilized for the preliminary testing as we had prior knowledge of their self-reported UE use from previous discussions. The preliminary testing was undertaken to: 1) identify category and/or question ambiguity; 2) identify potential difficulties with respondent bias and 3) perform member checking of experiences relayed in the initial focus groups.

### 3.2.1.5 Data analysis

Data analysis and data collection occurred concurrently which allowed for exploration of emerging themes in the subsequent focus group/interview sessions. Qualitative descriptive analysis (Sandelowski 2000) was used to analyse the content of the focus group transcripts. One of the common goals across all of the focus groups was to understand the descriptions of UE use in the language of healthcare providers and individuals with stroke. Qualitative descriptive analysis allows and encourages the researcher to stay true to the everyday language of the participants. Thus the researcher can make more surface level descriptions of the data without the requirement for deep level interpretation. In addition, the descriptive analysis method is particularly suitable for researchers who want to present the results in a straightforward manner that coincides with the questions asked.

Focus group transcripts within each participant group (ie. client, healthcare provider) were analysed separately and then as a whole to examine emerging patterns within an entire
participant group. Focus group findings were organized based on the types of questions presented to the groups. The findings from the healthcare provider focus groups were organized into three categories: 1) potential distinguishing factors of use and 2) potential use categories and 3) example tasks. The following information was extracted from the focus groups with individuals with stroke: 1) tasks performed and not performed by each person; 2) the role of the arm when performing the tasks and 3) descriptions of meaningful change. As hand dominance (in relation to side affected) and arm function/motor capability were significant factors related to UE use in the literature, the results of the client focus groups were examined by level of function/motor capability and side affected (i.e., affected dominant hand or affected non-dominant hand) to identify any emerging patterns.

### 3.2.1.6 Formation of the REACH scale

The REACH scale covered the full range of categories of UE use, from “no use” to “full use.” A conceptual definition of full use was created from information from the literature combined with focus group findings. This definition helped to guide the number and content of the other categories. Descriptions of each level used language and common examples provided from the clients when possible. Feedback was sought from the healthcare providers with regards to whether each category was clearly described and was clinically reasonable. Preliminary testing on individuals with stroke also ensured the scoring process was clear and straightforward. A checklist was added to each category to add clarity and ensure levels were mutually exclusive. The category checklists contained a set of criteria which had to be met in order for the respondent to be classified into that category. Finally, an algorithm that consisted of a series of yes or no questions was created to assist in the quick
classification of arm use level and help ensure low respondent and administrator burden. The
REACH version used in validity and inter-rater reliability testing is in Figure 3.1.

3.2.2 Evaluation of the REACH Scale

3.2.2.1 Design
Validity was tested using a cross-sectional hypothesis testing design. Inter-rater
reliability was assessed through independent administrations of the REACH scale at separate
times using different raters. All procedures were approved from the University of British
Columbia and Vancouver Coastal Health ethics boards.

3.2.2.2 Participants
Participants were recruited from the following sources: 1) former volunteers from the
research laboratory volunteer database; 2) former inpatients from a local rehabilitation
hospital within the last 5 years; 3) word of mouth via participants and 4) community stroke
recovery clubs. To be included in the study volunteers had to meet the following inclusion
criteria: 1) had a stroke at least 6 months prior to testing 2) experienced UE weakness
resulting from a stroke; 3) 19 years or older; 4) live in the community; 5) be able to
understand and follow instructions. Individuals were excluded from the study if they met any
of the following exclusion criteria: 1) have injuries in the muscles, bones or joints of either
UE that limited its function (e.g. fractures, severe arthritis); 2) have another neurological
condition in addition to the stroke; 3) unable to understand or communicate to the
investigators when asked questions. All participants gave informed written consent prior to
participating in the study.
3.2.2.3 Measures

The following section outlines the demographic information and post-stroke status measures collected from the participants during the testing session.

**Participant demographics:** Personal and stroke characteristics were collected to describe the sample. In addition, personal factors that were thought to influence UE use or the ability to reflect on UE use were also collected. The following characteristics were captured: age, gender, time post stroke, affected side, pre-stroke dominant side, living situation, education, attendance of inpatient rehabilitation program, stroke severity and cognition. Stroke severity and cognition were captured using the following clinical scales:

**National Institute of Health Stroke Scale (Appendix F):** The NIH stroke scale is a 15 item observational scale that captures common impairments following a stroke and is thus used as a composite measure of stroke severity (Brott et al. 1989). The scale assesses level of consciousness, vision, facial palsy, limb strength, ataxia, sensation, and speech and language. Evidence supporting the scale’s validity has been established (Lynden et al. 1991). Scores range from 0-42 with higher scores indicating greater stroke severity. The following cutoffs can be used to guide interpretation of stroke severity (Brott et al. 1989): ≥25 (severe neurological impairment); 5-14 (mild to moderate neurological impairment) and <5 (mild impairment).

**Montreal Cognitive Assessment (MoCA) (Appendix G):** The MoCA is a cognitive screening tool for the detection of cognitive impairment in older adults (Nasreddine et al. 2005). The scale captures attention, executive functions, memory, language, visuoconstructional skills,
abstract thinking, calculations, and orientation. The final score ranges from 0-30 where scores below 26 suggest mild cognitive impairment (Nasreddine et al. 2005). The scale has been validated in the stroke population (Pendlebury et al. 2010) however given its dependence on language, the scale is not considered valid for use with people with aphasia (Dong et al. 2010).

The following outcome measures of UE use, UE function and UE impairment were used in the validation of the REACH scale.

**Motor Activity Log-14-Amount of Use scale (Appendix H):** The Motor Activity Log-14 is a self-report measure that captures how much and how well the affected UE is used in 14 activities of daily living. Each activity is rated on separate scales (Amount of Use, Quality of Use) and the total composite score for each scale is an average of the activity scores. The total scores range from 0-5 with higher scores indicating greater amount or quality of use. Only the Amount of use scale was used in this study due to the high correlations between the Amount of Use and Quality of Use scales ($r=0.8-0.95$) (Van der Lee 2004; Uswatte et al. 2005; Uswatte et al. 2006b). Evidence to support the validity of the Motor Activity Log-14 has been reported (Uswatte et al. 2005).

**Accelerometers:** Actical™ accelerometers were used to capture affected UE activity in the home and community. Actical Aaccelerometers are light weight (17.5 grams), water proof monitors that are worn on the wrist. They capture UE activity by detecting acceleration in all 3 planes of movement. The monitors sample acceleration at 32 Hz and are sensitive to
acceleration from 0.05-2 G-force. The signal is processed as activity counts over a user-specified period or epoch. An epoch of 15 seconds was used in this study. The sensors capture both intensity and range of motion. Thus higher activity counts over a day can indicate higher movement speed, higher range of movement and/or higher duration of movement (Rand and Eng 2010). Evidence to support the validity of accelerometers in the stroke population has been reported (Gebruers et al. 2008).

**Action Research Arm Test (Appendix I):** The Action Research Arm Test is a 19-item performance-based measure of UE function for people with stroke. It assesses a client’s ability to handle objects of varying size, weight and shape (Platz et al. 2005). The test items are divided into 4 Guttman subscales: Grasp, Grip, Pinch and Gross Movement. The total score ranges from 0-57 with higher scores indicating greater functional ability. The Action Research Arm Test is widely used in the stroke literature and is often used as a comparison measure for validity testing of other UE functional measures (Baker et al. 2011). Evidence to support the scale’s validity has been reported (Baker et al. 2011).

**Stroke Impact Scale-hand subscale (Appendix J):** The Stroke Impact Scale is a self-report measure that captures consequences of stroke that are thought to impact one’s health (Duncan et al. 1999). The scale consists of 8 subscales that assess different consequences of stroke (e.g. mobility, communication, participation, hand function). The hand function subscale captures clients’ perceived ability to perform 5 different activities using their affected hands. The subscale score ranges from 0-100 with higher scores indicating higher perceived hand function. The validity of the 8 subscales has been investigated along with the
validity of the composite scale. Evidence to support the hand subscale of the Stroke Impact Scale has been reported (Duncan et al. 1999).

Chedoke-McMaster Stroke Assessment-Arm and Hand Recovery and Shoulder pain scales (Appendix K): The Chedoke-McMaster Stroke Assessment measures impairment and disability in clients with stroke. The measure consists of an impairment and function inventory. The impairment inventory assesses six domains. Three of these domains are arm recovery, hand recovery and shoulder pain. The arm and hand recovery domains are rated on a 7 point scale where 1 represents flaccid paralysis and 7 represents normal movement. Shoulder pain is also rated on a 7 point scale where 1 represents constant and severe shoulder pain and 7 represents absent shoulder pain and absent risk factors for shoulder pain. In this study, the arm and hand recovery scales were combined to form a composite arm and hand score. Thus scores range from 2 to 14 with 2 representing flaccid paralysis of the arm and hand and 14 representing normal movement of the arm and hand. Evidence to support the validity of the arm and hand composite scales and shoulder pain scales has been reported (Gowland et al. 1993).

3.2.2.4 Procedures
Volunteers who met the inclusion/exclusion criteria participated in one or two testing sessions. At the first testing session, the demographic information was collected and all the clinical measures (including the REACH scale) were administered. Administration of the tests was performed by two trained clinicians (Occupational therapist, Physical therapist). Participants willing to wear the accelerometers were provided with one monitor for each of their wrists. They were requested to wear the monitors during their waking hours for three
consecutive days starting the day after the monitors were provided. Participants were given instructions on how to wear the accelerometers properly and provided with a take-home pamphlet which reiterated the instructions and included picture examples of the correct and incorrect orientations of the monitors. Although the monitors are water proof, participants were told they could remove the monitors for bathing if they were worried about discomfort from wearing a wet strap. Participants were advised not to wear the monitors in the pool.

All participants were invited to participate in the optional second testing session which provided data for the interrater reliability analysis. The latter session only involved the second administration of the REACH scale. The two trained clinicians administered the REACH scale. The rater who administered the REACH at the first testing session was different from the rater who administered the REACH at the second testing session. The mean time interval between testing sessions was 7 days (range: 3-21 days).

3.2.2.4 Data analysis

Descriptive statistics were used to examine the participant demographic and clinical characteristics. Distribution of the data was assessed for normality using the Shapiro Wilk test. We proposed a sample size of 90 to capture an average of 15 subjects in each of the 6 categories, with a minimal representation of 5 subjects in any one category. In order to test the hypothesis that the REACH scales would have correlations of at least 0.50 with the selected outcome measures, a sample size of 30 participants per scale was required (\(\alpha=0.05; \beta=0.20\)) (Portney and Watkins 2009). Using calculations derived by Donner and Eliasziw (1987) where \(\alpha=0.05\) and \(\beta=0.20\), a sample size of 12 per group is required in order to test the hypothesis that the ICC exceeds a value of 0.80 with a predicted “true” ICC value of
SPSS version 18.0 was used to calculate the ICC values and Stata 11.0 was used to calculate the weighted kappas.

**Reliability**

Interrater reliability was examined using the ICC and weighted kappa with linear weights. A two-way random effects model with absolute agreement was selected for calculation of the ICC. The two-way random model was selected under the assumption that the two raters used in this study represent a random selection of all possible raters. The two-way model, which allows us to isolate the error introduced by the raters, is also appropriate in this case as only two raters were used to assess reliability. This means that each rater assessed the same number of participants in the reliability sample (Shrout and Fleiss, 1979). Linear kappa weights were selected due to their reduced dependence on the number of scale categories (Brenner and Kliebsch 1996). In addition, linear weights were selected because of the observation that weighted kappas calculated using quadratic weights are often equivalent to the ICC (Brenner and Kliebsch 1996; Streiner and Norman 2008; Portney and Watkins 2009). The standard error of measurement (SEM) and the minimal detectable change (MDC) calculated at 95% confidence level were used to examine the stability of the REACH scale. This was possible due to the use of a time interval in between the testing sessions. The SEM and MDC\textsubscript{95} were used to provide a preliminary estimate of meaningful change. Although the literature recommends that these values not be used in isolation to represent important change (Revicki et al. 2008), they do provide information about the tool’s ability to capture change (Beaton et al. 2001). For instance, confidence that true functional change has been observed in a study is increased when the observed change is equal to or surpasses a measure’s SEM or MDC\textsubscript{95} values (Beaton et al. 2001).
Validity

Hypothesis testing was used to examine the construct validity of the REACH scale. Expected correlations between the REACH scale scores and other measures of UE use, UE function and UE impairment formed the basis of the hypotheses. These hypotheses were generated after considering the construct of use captured by the REACH scale. The hypotheses generated were: (1) The REACH scale will have a positive and strong relationship to the Motor Activity Log, Action Research Arm Test, Stroke Impact Scale-hand scale, and Chedoke-McMaster Arm and Hand Scale (ρ>0.70). (2) The REACH scale score will have a positive and medium relationship to affected UE activity counts and the Chedoke-McMaster pain scale (ρ=0.30-0.70). Correlation coefficient values were selected based on cutoffs adapted from Portney and Watkins (2009). The Spearman rank correlation coefficient was selected due to the ordinal nature of the REACH scale. Scatterplots of the REACH scores and the measures used in the hypotheses were also generated in order to visually examine the nature of the relationships. All validity analyses were conducted using SPSS version 18.0.

Participants who agreed to wear the accelerometers were asked when the monitors were worn and removed during the time they were in the participants’ possession. Individuals were excluded from the analyses if they did not wear the accelerometers for long periods (>5 hours) during any three consecutive days. Upon return of the accelerometers, the graphs of the activity counts were visually examined. To account for differences in wake and sleep times, three 24 hour periods were examined and the average daily activity counts of the affected UE was calculated. The amount of activity was very minimal when some individuals wore the monitors when they were sleeping. Participants associated with accelerometers that
demonstrated irregular and/or unexplainable activity were also excluded. Examples included constant large spikes throughout the 24 hour period, very low levels of activity in both arms (<3000 counts), large random spikes indicative of someone waving their arms very quickly or large differences between the total wearing time of the affected and non-affected UEs. One participant was excluded from the analysis as the accelerometer could not be read upon its return. A total of 78 individuals agreed to wear the accelerometers. After the above participants were excluded, a total of 68 individuals remained in the analysis.

3.3 Results

3.3.1 Focus group participants

A total of 33 participants (13 clinicians, 16 participants with stroke, 4 caregivers) took part in the focus groups/interviews. Table 3.1 outlines the characteristics of the participants. Healthcare providers from 5 disciplines were represented in the focus groups with occupational therapists making up the highest proportion. The clinicians had an average of 16 years of experience (range 1.3-34) where the majority of those years were spent working with the stroke population. Individuals with stroke were several years post stroke (mean of 7.8) and the number of participants with no active movement, proximal movement only and proximal and distal movement were almost equal. Finally, just over half of the participants with stroke had their non-dominant UE affected.

3.3.2 Focus groups findings

3.3.2.1 Healthcare providers

Healthcare providers were asked to identify factors that could distinguish between different levels of affected UE use. Participants identified the following factors: 1) role of the
UE during an activity (e.g. stabilizer, reach and grasp, “doer or dominant hand”); 2) amount of assistance required to use the UE; 3) frequency of use; 4) types of tasks people are performing using the UE (e.g. strength, gross motor, fine motor/dexterity, tasks that require precision); 5) quality of movement (speed, accuracy); 6) types of circumstances when the UE is used (e.g. unilateral activities, bilateral activities, when unaffected UE is occupied) and 7) use of compensatory strategies (e.g. adaptive equipment, “switching arms”). Many participants also discussed the factors that explain use. The following factors were identified: spasticity, pain, proximal vs distal control, confidence or trust in the ability of the UE.

Participants identified many different activities throughout the group discussions. The majority were activities of daily living in the home setting. The tasks varied from simple (requiring gross motor control) to complex (requiring fine motor control), although the majority of the activities identified involved use of the hand. Examples of activities mentioned that require gross motor control were: stabilizing a piece of paper, pushing an item forward, holding a grocery bag, pushing a door open, and washing arm using a washcloth. Examples of activities mentioned that require more fine motor control include: shaving, cleaning self following toileting, using a debit machine, doing up buttons/zippers, typing and gesturing. Finally, the majority of participants believed that 5 levels would capture the full range of use from no use to full use and provide enough distinctions in UE use. A question brought up by many clinicians concerned the best way to define full use. The majority of clinicians felt that this was an important aspect of the scale that required a clear definition.
3.3.2.2 Clients and caregivers

Individuals with stroke and their caregivers were asked to describe how they used or didn’t use their affected UE in their everyday lives. Participants provided a rich description of their use by identifying specific tasks and describing how the UE was involved during those tasks. Participants also identified multiple strategies they used to compensate for their functional limitations. Although each participant identified specific tasks that were not always mentioned by other participants, patterns of the types of tasks and the manner in which the affected UE was used did emerge. For instance, the majority of participants with limited hand function (i.e. could not actively open their hand) identified using their affected UE to stabilize objects regardless of whether their dominant or non-dominant side was affected. Beyond stabilization, there was quite a large variation in the types and amounts of activities individuals with limited hand function were performing. These examples were unique to the types of tasks brought up in the clinician focus groups. Clients described using the affected hand as a passive clamp, using the affected arm to flip a light switch or using the affected arm to feed clothes through a sewing machine.

Hand dominance appeared to play a larger role among participants with greater functional capacity. Participants with hand function (i.e., could actively open hand) and their dominant hand affected identified situations when compensations were used. Compensation was mainly defined as either using the non-affected UE entirely or using the affected UE as an assist. Compensations were primarily used when use of the affected UE would result in undesirable consequences or if the affected UE did not have the strength, coordination or range of motion required to perform the task. Participants who had their non-dominant UE affected and possessed hand function mainly described their use in the context of performing
bilateral activities. Finally, participants expressed difficulties with describing what a meaningful but small improvement in their UE use would look like. This appeared to be especially difficult for individuals with limited arm and hand function (cannot actively move shoulder or open the hand) and very high function (able to use the hand). Ultimately, improvement for many participants meant the ability to perform tasks they no longer were able to perform. The examples were unique across participants. Examples of tasks identified included: having more choice in the types of clothes they wear, being able to use small tools, wearing shoes with shoelaces, stabilizing objects with the hand, using the UE for anything at all and going swimming. Participants appeared to find it easier to describe what improvements in their UE would be necessary for them to use their affected UE more. Necessary improvements included: greater range of movement at the shoulder, any use of the hand, any movement at all, increased dexterity and coordination of the hand, increased strength, decreased spasticity and increased sensation.

In summary, participants with stroke provided many examples of how their affected UE was used in the context of tasks they were able or not able to perform. These findings provided the following information for the development of the REACH scale: 1) examples of tasks performed or not performed by the affected UE; 2) manner in which the affected UE was used (e.g. as a stabilizer, assist); 3) circumstances when the affected UE wasn’t used (e.g. undesirable consequences, limited function); 4) greater distinctions at the lower end of function (ie. limited hand function) than originally considered; 5) differences in descriptions in use based on concordance of affected and dominant sides and; 6) the importance of including example tasks in the scale to help cue future scale respondents about their UE use.
3.3.3 The REACH scale

In conceptualizing typical upper limb use, the following assumptions were adopted from Barreca et al. (2003): 1) efficient performance of daily home and community activities involves the cooperative use of both upper limbs; 2) both upper limbs fulfill the following roles during typical performance of daily activities: reach and grasp, stabilize and manipulate objects and 3) typical use of the upper limbs involves an interaction with objects of various sizes, weights and locations. This conceptualization was used to help guide the design of distinct categories.

Also, in typical upper limb use, the dominant and non-dominant sides often perform unique roles. For instance, during bilateral tasks, the dominant hand usually performs the manipulation action while the non-dominant hand often acts as the support. Separate scales were therefore developed. In addition, focus group/interview responses from individuals with stroke helped to evolve the scale into 6 distinct categories. In particular a level was added to reflect the wide variation in tasks reported by participants with limited hand function. Thus the REACH scale consists of two separate classification scales for people who had their dominant side and non-dominant side affected by the stroke. Both scales classified use into 6 levels that range from “no use” to “full use”. Distinctions between levels are based on the primary roles of the affected UE during everyday activities, the types of activities the affected UE performs, extent of use and use of adaptive strategies. Full use is defined relative to the client’s use prior to the stroke.

The REACH scales consist of an algorithm and a checklist (Figure 3.1). The algorithm can be used to quickly narrow down a client’s appropriate classification level. The checklist should then be used to verify the level determined by the algorithm. The checklist
3.3.4 Evaluation of the REACH scale

3.3.4.1 Sample
A total of 96 individuals participated in the evaluation component of the REACH scale. Table 3.2 outlines the demographic and clinical characteristics of the sample. In summary participants were drawn from a wide range of age groups (32-96) with a relatively young mean age of 64.4. Men and women were represented equally in the sample and just over half of the sample had their dominant sides affected (54%). The majority of the sample lived with a spouse or family member and almost all participants had attended an inpatient rehabilitation program. Finally, the participants represented individuals with mild to moderate stroke severity as captured by the National Institute of Health Stroke Scale.

3.3.4.2 Distribution of the REACH scale
Table 3.3 displays the number of participants classified into the REACH categories. For both the non-dominant and dominant affected groups, the greatest number of individuals used their affected UEs to perform some reach and grasp tasks that require hand manipulation (Level 3). The next highest proportion of people used their affected UEs to stabilize objects (Level 1) or used it for everyday tasks except when there were potential negative consequences (Level 4). The mean REACH sample score was very similar across the non-dominant (2.6) and dominant affected groups (2.3) (Table 3.3). Both the visual examination of the REACH scores and the Shapiro Wilk test (Table 3.4) indicated a non-normal distribution of scores. The average time to administer the REACH scale was 5 minutes.
3.3.4.3 Reliability
A total of 73 participants agreed to attend the second testing session and were included in the reliability analyses. Scatterplots of the data demonstrated that the dominant and non-dominant affected subjects had similar relationships between Rater 1 and Rater 2 with similar slopes and range of data (Figure 3.2). Thus, all participants (dominant and non-dominant affected subjects) were collapsed together and the ICC (2,1) for the REACH scale was 0.968 (95% CI: 0.949-0.980, p<0.001) and the $\kappa_w$ was 0.908 (95% CI: 0.893-0.926, p<0.001). The SEM was 0.27 and the MDC$_{95}$ was 0.75. These values represent 11% and 31% of the mean REACH score respectively. Table 3.4 displays the reliability coefficients, SEM and MDC$_{95}$ values obtained for the REACH scale. Sub-analysis justified the collapsing of the data as the ICC and weighted kappa ($\kappa_w$) coefficients for the dominant and non-dominant affected scales differed by less than 0.05.

3.3.4.4 Validity
Scatterplots between the REACH scale and other variables demonstrated that the dominant and non-dominant affected subjects had similar relationships with similar slopes and range of data (Figures 3.3A-F), and thus, the data were collapsed together. Table 3.6 displays the mean scores for the outcome measures used in the reliability analyses.

Except for the relationship between the REACH scale and shoulder pain, the correlations were in the direction and within the range of magnitudes hypothesized. Table 3.5 displays the calculated correlation coefficients and Figure 3.3 (A-F) displays the scatterplots of the relationships. As expected, the REACH scale was strongly correlated to the Motor Activity Log, Action Research Arm test, Stroke Impact Scale and the Chedoke-McMaster-Arm and Hand scales ($\rho$=0.87-0.935, p<0.001). Although in the expected direction, the
relationship between the REACH scores and shoulder pain (as captured by the Chedoke-McMaster shoulder pain scale) were lower than expected (observed $\rho=0.24$ vs expected $\rho=0.30-0.70$). Although the correlation between the REACH and shoulder pain obtained significance ($p=0.02$), examination of the scatterplot suggests little relationship between the REACH scale and shoulder pain exists at all (Figure 3.3F). Inspection of the scatterplots also revealed interesting differences between the distribution of the REACH scale scores in relation to the Motor Activity Log, the Action Research Arm test and the Chedoke-McMaster Arm and Hand scale (Figure 3.3A,C,E). These graphs show there is greater distribution across REACH scores at the lower ends and higher ends of use, resulting in a slight curvilinear “S” relationship between the REACH scores and these impairment and functional scales. Sub-analysis justified the collapsing of the data as the differences in the correlation coefficients between the dominant and non-dominant hand scales showed very small differences ($\rho<0.03$), except for shoulder pain ($\rho=0.18$).

3.4 Discussion

Capturing outcomes that are meaningful to people with stroke are becoming increasingly important as the demand for accountability in stroke care rises (Reeves et al. 2010). Moreover, a client’s activity outside of a controlled environment is being recognized as an important outcome in stroke rehabilitation (Lang et al. 2012; Dobkin et al. 2011; Baker et al. 2011). The development of the REACH scale introduces an exciting option for assessing upper extremity activity in a person’s own environment (home and community), and for discriminating between different levels of UE use which are meaningful to
individuals with stroke. The following section discusses both the development of the REACH scale and the results of the evaluation of this measure.

### 3.4.1 Development of the REACH scale

The development process employed in this study led to the creation of two separate scales which consider whether the dominant or non-dominant hand has been affected. Each scale consisted of 6 levels that ranged from no use to full use. Iterative feedback from clinicians, clients, and caregivers, combined with preliminary testing on individuals with stroke led to the final version of the REACH scale (Figure 3.1).

Focus groups with clinicians, clients and caregivers were employed in order to better understand upper extremity use from their perspective and to guide the generation of the number and content of the scale categories. While the language and organization of the discussions varied across the different focus groups, it was striking that a number of common themes arose across the clinician and client groups. Importantly, both groups were able to identify distinct roles in how the affected UE is used and these salient descriptions of use formed the basis of the scale content. More specifically, components of all of these themes were included in the scale to help distinguish between different levels of use.

Analysis of the clinician and client focus group discussions also revealed important differences between the two groups. The most evident difference related to the number of categories required to provide meaningful distinctions in UE use. For instance, the clinicians believed that 5 levels would provide enough differentiation between levels of use when capturing no use to full use. Clients at lower levels of UE function (ie. limited ability to open hand) however articulated a wide variation in the types of tasks they were performing. This finding supported the addition of an extra level of use at the lower end of the scale, which
resulted in the creation of a 6 level scale. Another important difference between the clinician and client group findings was the manner in which frequency of use was understood. The majority of clinicians believed frequency of use was an important factor in the distinction between different levels of use. Many of the clinicians proposed using percentages to distinguish use as this represented a clear cutoff. Although some stroke clients were able to attach a number to their amount of use (i.e., 50% less than before the stroke), others found the task of quantifying their use very difficult. Indeed, upon preliminary testing of the scale, the use of percentages to distinguish between amounts of use identified a potential source of confusion for the clients and the administrators of the test. For instance, responses from stroke clients to the question regarding frequency of use included comments such as “my arm is about 50% of what it was before” or I have “75% of the strength back”. It was therefore difficult to identify whether stroke clients were imagining their pre-stroke amount of use or their pre-stroke function. Ultimately this finding resulted in the removal of percentages to capture amount of use.

The final important feature of the REACH scale that was guided by both the feedback and literature was the creation of separate scales for clients with their dominant side or non-dominant side affected. Differences between dominant and non-dominant affected side clients were prominent in the clients’ descriptions of use. This was especially observed among clients with greater UE function. Clients with their dominant side affected and non-dominant side affected had different expectations for their affected UE use and described the manner in which they used their UE differently (e.g., the doer vs the assist). The types of activities used as examples were also different between dominant and non-dominant affected participants (e.g., unilateral vs bilateral activities). An association between UE use and
dominant affected or non-affected individuals was also identified in the literature (Haaland et al. 2012; Rinehart et al. 2009; Uswatte et al. 2006a). In particular, these studies observed different patterns of use among dominant and non-dominant affected individuals. Creating separate scales allowed us to account for differences in the patterns and descriptions of use while allowing us to capture a similar progression from no use to full use.

Finally, we did not observe an even or normal distribution in the REACH dominant or non-dominant side affected scores in our sample. This observation is not surprising given the chronic nature of the participants in our sample (mean of 7 years post stroke). The scales were developed to capture the full spectrum of use from no use to full use. If we collected the REACH scores from the time of stroke through to the chronic phase, we would expect people to move through these different categories.

3.4.2 Evaluation of the REACH scale

The evaluation of the REACH scale’s interrater reliability and validity revealed important findings about two major measurement properties of the scale.

3.4.2.1 Reliability

The reliability analysis of the REACH Scale obtained an ICC of 0.97 and a linear weighted kappa of 0.91. Current measurement guidelines do not include strict standards for the acceptable magnitude of an ICC or kappa coefficient (Mokkink et al. 2010; Streiner and Norman 2008; Portney and Watkins 2009). Streiner and Norman (2008) and Portney and Watkins (2009) recommend researchers or clinicians using the tool should make the judgment concerning the acceptable magnitude and precision of the reliability coefficient.

This study uses both an ICC and a weighted kappa coefficient to capture the interrater reliability of the REACH scale. Controversy in the literature concerning the most appropriate
Coefficient to use with an ordinal scale was the main reason for adopting both coefficients (Connell and Tyson 2012; Mokkink et al. 2010a; Streiner and Norman 2008). The choice between these two coefficients is especially difficult when the scale has an intermediate number of categories such as 6. A kappa coefficient is more appropriate for scales with few categories (Brenner and Kliebsch 1996). Accordingly the fact that the kappa coefficient is smaller than the ICC is consistent with the observation that the kappa with linear weights decreases as the number of categories increase (Brenner and Kliebsch 1996). An ICC of 0.97 with a lower bound of 0.95 and a weighted kappa of 0.91 with a lower bound of 0.89 both surpass guideline values that indicate good reliability however (Portney and Watkins 2009; Streiner and Norman 2008).

The reliability design, which involved different raters administering the scale at two different time points (separated by 1 week), was selected to minimize the chance of administrator and participant memory bias and to reflect a realistic clinical scenario. In the practice setting, it is very common for different clinicians to perform baseline and follow up assessments. The obtained values for interrater reliability are thus conservative estimates as they capture both the effect of different raters and the effect of time.

Finally, the calculated SEM and MDC$_{95}$ values provide preliminary estimates of meaningful change in the REACH score. With an SEM value of 0.27 and a MDC$_{95}$ value of 0.75, a change in one category is likely a useful indicator of real or meaningful change. Future work that incorporates clients’ perceptions of meaningful change in REACH scores is necessary in order to confirm this finding.
### 3.4.2.2 Validity

Hypothesis testing was used to examine the validity of the REACH scale. Except for the expected relationship between the REACH and Chedoke McMaster-shoulder pain scores, the results are consistent with all of our hypotheses. We expected relationships between the REACH scores and measures of UE use, function and impairments. These hypotheses were generated in consideration of the construct of use captured by the REACH scale. Thus, confirmation of the majority of our hypotheses provides evidence to support the validity of the REACH scale.

We observed a very strong relationship between the REACH and the Motor Activity Log scales ($\rho=0.94$) and a moderate relationship between the REACH and affected UE activity counts ($\rho=0.61$). A stronger relationship with the Motor Activity Log was expected as activity counts are a measure of both functional and non-functional activity. The Motor Activity Log and the REACH scales both capture functional activity of the affected UE. The observed correlation between the REACH and activity counts is within the range of correlations (0.47 to 0.74) cited between the Motor Activity Log and accelerometry (Gebruer et al. 2008). The recommended time for the Motor Activity Log-14 and the Motor Activity Log-30 is 30 and 60 minutes respectively (Taub et al. 2011). With the average administration time of 5 minutes, and a high correlation to the Motor Activity Log, the REACH scale represents a promising alternative self-report tool for categorizing UE use. Interestingly, examination of the scatterplots of the REACH scale with measures of use, function and impairment (Figure 3. 3.3A,C,E) suggested a curvilinear relationship between the REACH and some of the measures of use, function and impairment. Greater variation in REACH
scores were observed at the lower and higher ends of these other measures. This finding may suggest a better ability to discriminate between people at lower and higher values of use.

Strong relationships were also observed between the REACH and measures of UE function ($\rho=0.93-0.94$) and arm and hand motor impairment ($\rho=0.91$). Although consistent with our hypotheses, these relationships are stronger than the relationships observed in the literature. Cited correlation coefficients between UE function and UE use (captured by the Motor Activity Log and accelerometry) range from 0.40-0.82 (Chen et al. 2012; Michielson et al. 2009; Harris et al. 2007; Lang et al. 2007; Dromerick et al. 2006; Uswatte et al. 2006b; Van der Lee et al. 2004; Van der Pas et al. 2001). Correlation coefficients between UE motor impairment and UE use (captured by the Motor Activity Log and accelerometry) range from 0.54-0.85 (Thrane et al. 2011; Michielson et al. 2009; Gebruers et al. 2008). One possible explanation for the strong correlations is that progression on the REACH scale from a lower category to a higher category relates to the completion of tasks that require greater motor capacity and function.

Our findings did not support the hypothesis that the REACH score would have a moderate relationship to the Chedoke-McMaster Shoulder pain scale. Examination of the scatterplots revealed no relationship between these two scales whereas the correlation analysis indicated a weak but significant positive relationship. One explanation for these findings is the large number of people in our study with no or little shoulder pain. Alternatively the lack of an observed relationship could be due to inclusion of tasks (e.g., hand tasks) in the REACH scale that can be performed even when pain is present. The association between UE pain and UE use is not well studied. Only two papers examined this relationship (Harris and Eng 2007; Lang et al. 2007), and only one of these studies found a
significant relationship (Lang et al. 2007). Future studies that include a larger number of people with shoulder pain are needed to confirm this apparent lack of relationship between UE use and shoulder pain.

Finally, the observation that the dominant side and non-dominant side affected scales had similar relationships to the external measures provides further evidence of the validity of the REACH scale. Separate scales were created due to the differences in use described by the dominant affected and non-dominant affected participants in the focus groups. The two scales are able to account for these differences by including unique descriptions of use within each category. In development of these two scales however, it was our goal for both scales to provide a similar progression from no use to full use. The similarity of the non-dominant and dominant scales to the external measures suggests we achieved our goal of capturing a similar progression.

3.4.2.3 Limitations
The mean age of participants in this study was 64. Compared to the stroke population as a whole, study participants were relatively younger. It is possible that older individuals may use their affected UEs differently than their younger counterparts. In addition, we did not observe an even distribution of REACH scores across the six levels. In particular, there were relatively few participants classified at level 2. This result may be a product of sampling variation or may be indicative of the true distribution of participants in the population. In addition, individuals may move through the lower stages at more acute times post stroke. Future studies that examine the REACH score over time can shed light on the relationship between time post stroke and REACH score. Longitudinal studies would also be able to address the responsiveness of the REACH scale.
3.5 Conclusion

This study outlined the development and evaluation of a new classification scale that captures UE use in the home and community setting (REACH scale). Focus groups with clinicians and clients revealed common themes and unique perspectives that were incorporated into the content and structure of the measure. Feedback from clinicians and preliminary testing on individuals with stroke further refined the clarity and content of the scale. Evaluation of the interrater reliability and validity of the REACH scale provide compelling evidence regarding these measurement properties. The results presented here suggest the REACH scale offers an efficient and effective way to capture UE use in clients’ community and home settings.
4. Conclusion

This thesis described two studies undertaken to develop a new classification scale that captures UE use following stroke (Chapter 2 and 3). More specifically, the goal was to create an ordinal scale that provides meaningful descriptions of, and distinctions between, different levels of UE use. Chapter 2 confirmed the lack of existing measures of UE use in the home and community setting. This systematic review also revealed that the existing literature provides little information about the characteristics and content of meaningful use of the UE. In light of this gap, it was necessary to seek insights from individuals with stroke and clinicians who work with stroke clients.

Chapter 3 employed a focus group methodology to guide content development for the new scale. Construction of the new scale also incorporated clinimetric principles. These principles encourage researchers to develop scales meeting the standards of clinical sensibility. The process adopted in creating this scale was designed to enhance the clinical sensibility of the measure. Key phases of the process aimed at ensuring such sensibility include: consultation with clinicians and clients on their perceptions of meaningful UE use; refinement of the scale after seeking further feedback from clinicians; and a preliminary testing stage involving individuals with stroke. Furthermore, these steps helped ensure the resulting tool has the following characteristics: clarity in the administration process; a well-defined description of distinct categories; an acceptable scope; and a tool that can be administered in an efficient and straightforward fashion.

This tool development process resulted in the Rating of Everyday Arm-use in the Community and Home (REACH) scale. The REACH measure involves two related scales that account for differences in use among dominant and non-dominant affected individuals.
The tool also includes an algorithm, consisting of yes or no questions, that both fosters efficiency in categorizing clients and increases the clarity of distinctions between levels.

Analysis of the evaluation data revealed that the two REACH scales had similar reliability coefficients and similar correlations with external measures. These results indicate both scales capture a similar progression of UE use from no use to full use. Finally, high interrater reliability coefficients and clear support for a range of validity hypotheses provide strong support for these measurement properties.

Important additional insights can be gained from future research that incorporates the REACH scale in a longitudinal study that assesses change in UE use over time. Further studies might also assess the ability of the REACH scale to capture meaningful change in UE in the home and community settings. The statistical estimate of important change calculated in Chapter 3 (captured as a MDC$_{95}$) suggests that movement of one category is meaningful. This is similar to the MDC$_{95}$ values calculated in Chapter 2 for the Motor Activity Log (0.67-1.27) (Appendix D). Notably, both the Motor Activity Log and the REACH scale range from 0-5. Moreover, while the progression of levels in the REACH scale is informed by participants’ perceptions of meaningful change, a longitudinal study designed to capture important change is warranted.

A related question concerns whether treatments can result in a change of one category in the REACH scale. Moving an entire category involves substantial change in use. For instance, a change from level 1 to level 2 requires a shift from using the arm or hand as a passive clamp to using the arm actively to reach out for something that requires minimal hand manipulation (flipping a light switch). Future research can assess whether current treatments could exert such change.
In summary, several novel findings were obtained through the process of completing this thesis. Chapter 2 revealed key findings related to the ability of outcome measures that capture UE function following stroke to capture important, observed and detectable change. We found that the magnitude of the important change value can vary largely depending on the method of calculation used. We also observed that rehabilitation treatments can change patient-perceptions of function as effectively as performance-based measures of function. Finally the results from Chapter 2 also suggest that larger sample sizes may be necessary in order to detect true change in patient-perceived UE function given the larger measurement error associated with these types of tools.

Through the process of developing the REACH scale, we observed agreement between clients and clinicians’ descriptions of UE use. In addition, clients were able to articulate their perceptions of meaningful change in UE use using task-oriented examples. Finally, the multistage process used to develop the REACH scale was both feasible and effective. The development process resulted in the creation of a tool that distinguishes between different levels of affected UE use following stroke and evaluation of this tool revealed evidence supporting its interrater reliability and validity.
### Tables

**Table 1.1 Relevant questions for assessment of clinical sensibility**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Relevant Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>Has justification of the scale been provided? What settings is the scale suitable for? What was the intended function of the scale (description, evaluation, prediction)</td>
</tr>
<tr>
<td>Overt Format</td>
<td>Is the format of the scale easy to understand? Is the output scale transparent? Are the instructions for administration clear and thorough?</td>
</tr>
<tr>
<td>Suitability of output scale</td>
<td>Are the categories mutually exclusive? Are the categories realistic? Does the scale have a comprehensive scope?</td>
</tr>
<tr>
<td>Face Validity</td>
<td>Are questions designed to elicit accurate and informative answer? Do the scale items reflect basic evidence? Are the combinations of the components of the scale clinically possible?</td>
</tr>
<tr>
<td>Content Validity</td>
<td>Are there any important content omissions? Are there any inappropriate content inclusions? Is there appropriate weighting of the scale’s components?</td>
</tr>
<tr>
<td>Ease of use</td>
<td>Is the scale easy to use? Is the scale quick to administer? Is any equipment necessary? How much training is required to administer the scale?</td>
</tr>
</tbody>
</table>

Adapted from Feinstein (1987)
### Table 1.2 Five aspects of construct validity and their associated sources of evidence

<table>
<thead>
<tr>
<th>Validity Aspects</th>
<th>Sources of evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>Content sources of evidence are concerned with the relevance and representativeness of the scale components in relation to the construct</td>
</tr>
<tr>
<td>Substantive</td>
<td>Substantive sources of evidence are concerned with the processes underlying the test takers responses (i.e. are test takers interpreting the questions in the manner that was intended)</td>
</tr>
<tr>
<td>Structural</td>
<td>Structural sources of evidence are concerned with the relationship between test items or the dimensionality of the scale and whether it coincides with the dimensionality of the construct</td>
</tr>
<tr>
<td>External</td>
<td>The external sources of evidence are concerned with whether the test scores are related to external test scores in the manner that coincides with the construct theory</td>
</tr>
<tr>
<td>Consequential</td>
<td>The consequential sources of evidence are concerned with the appropriate positive and negative consequences of the test results</td>
</tr>
</tbody>
</table>

Adapted from Messick (1989)
Table 2.1 Number of articles for included measures by category of change

<table>
<thead>
<tr>
<th>Measure</th>
<th>Important Change</th>
<th>Detectable Change</th>
<th>Observed Change (Natural Recovery)</th>
<th>Observed Change (due to Treatment)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARAT</td>
<td>1</td>
<td>2</td>
<td>14</td>
<td>10</td>
<td>27</td>
</tr>
<tr>
<td>MAL</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>23</td>
<td>29</td>
</tr>
<tr>
<td>Wolf</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>12</td>
<td>20</td>
</tr>
<tr>
<td>SIS</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Frenchay</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>CAHAI</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>ABILHAND</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Jebsen</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Accelerometry</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>TEMPA</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>FTHUE</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>AMAT</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Duruoz</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>HFS</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

ARAT: Action Research Arm Test; MAL: Motor Activity Log; Wolf: Wolf Motor Ability Test; SIS: Stroke Impact Scale; Frenchay: Frenchay Arm Test; CAHAI: Chedoke Arm and Hand Activity Inventory; Jebsen: Jebsen Hand Function Test; TEMPA: Upper Extremity Performance Test for the Elderly; FTHUE: Functional Test for the Hemiplegic Upper Extremity; AMAT: Arm Motor Ability Test; Duruoz: Duruoz Hand Index; HFS: Hand Function Survey
Table 2.2 Minimally clinically important difference (MCID) values calculated by distribution and anchor-based methods

<table>
<thead>
<tr>
<th>ABILHAND</th>
<th>Accel</th>
<th>ARAT</th>
<th>MAL</th>
<th>MAL</th>
<th>SIS</th>
<th>Wolf</th>
<th>Wolf</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QOM</td>
<td>AOU</td>
<td>FAS</td>
<td>Time</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**MCID obtained by distribution-based methods**

| Effect size | 0.26\(^1\) (in logits) |       | 5.8\(^4\) (5.8%) | 0.14\(^3\) (2.8%) | 1.4\(^5\) (1.1%) |

**MICD obtained by anchor-based methods**

<table>
<thead>
<tr>
<th>% Recovery</th>
<th>0.35(^1) (in logits)</th>
<th></th>
<th></th>
<th>0.33(^3) (6.6%)</th>
<th>1.64(^5) (1.4%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>NS(^2) 12-17(^2) (21-30%)</td>
<td></td>
<td>1.0-1.1(^2) (20-22%)</td>
<td>1.0-1.2(^2) (20-24%)</td>
<td>19.0(^2) (16%)</td>
</tr>
</tbody>
</table>

\(^a\) MCID values are displayed in raw scores (% of scale maximum score)

\(^b\) An effect size of 0.2 was used as the distribution method to determine MCID

\(^c\) Anchor-based methods used to calculate MCID consisted of: Fugl-Meyer, change of 6-10 points on the Fugl-Meyer upper extremity scale; % Recovery, 10-15% or 50% recovery on the Stroke Impact Scale global recovery item; Global, perception of important change on a global rating scale

Accel: Accelerometry; ARAT: Action Research Arm Test; MAL: Motor Activity Log; QOM: quality of movement scale; AOU: amount of use scale; SIS: Stroke Impact Scale; Wolf: Wolf Motor Function Test; FAS: functional ability scale; NS: MCID values could not be calculated due to no relationship between change scores and global recovery scales; logits: log odds scale units that allow Likert scale scores (i.e. ABILHAND raw scores) to be interpreted as interval scores (Streiner and Norman 2008)

\(^1\) Wang et al. 2011; \(^2\) Lang et al. 2008; \(^3\) Fritz et al. 2007; \(^4\) Lin et al. 2010; \(^5\) Lin et al. 2009b
Table 3.1 Focus group participant characteristics

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Individuals with Stroke</th>
<th>Caregivers</th>
<th>Healthcare Providers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=16</td>
<td>N=4</td>
<td>N=13</td>
</tr>
<tr>
<td>Female, N (%)</td>
<td>9 (56%)</td>
<td>2 (50%)</td>
<td>13 (100%)</td>
</tr>
<tr>
<td>Age, mean (SD)</td>
<td>65.3 (7.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years post stroke (SD)</td>
<td>7.8 (3.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dominant hand affected, N (%)</td>
<td>7 (44%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AHA stroke functional classification, N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>10 (63%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>2 (12%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>3 (19%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>1 (6%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active Range of Motion, N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>4 (25%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoulder only</td>
<td>5 (31%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoulder and hand</td>
<td>7 (44%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years practicing, mean (SD)</td>
<td></td>
<td>16.3 (9.7)</td>
<td></td>
</tr>
<tr>
<td>Profession (N)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupational Therapist</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physiotherapist</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychiatrist</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nurse</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recreational Therapist</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AHA: American Heart Association
Table 3.2 REACH evaluation participant characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Results (N=96)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean (SD; range)</td>
<td>64.4 (11.7; 32-96)</td>
</tr>
<tr>
<td>Females, N (%)</td>
<td>47 (49.0%)</td>
</tr>
<tr>
<td>Years post stroke, mean (SD; range)</td>
<td>7.0 (5.4; 0.5-23.3)</td>
</tr>
<tr>
<td>Right side affected, N (%)</td>
<td>52 (54.2%)</td>
</tr>
<tr>
<td>Dominant Side Affected, N (%)</td>
<td>52 (54.2%)</td>
</tr>
<tr>
<td>Living Situation, N (%)</td>
<td></td>
</tr>
<tr>
<td>Spouse/Family</td>
<td>62 (64.6%)</td>
</tr>
<tr>
<td>Alone</td>
<td>29 (30.2%)</td>
</tr>
<tr>
<td>Assisted Living</td>
<td>5 (5.2%)</td>
</tr>
<tr>
<td>Years of Education, mean (SD; range)</td>
<td>14.7 (2.9;4-21)</td>
</tr>
<tr>
<td>Attended Inpatient Rehabilitation, N (%)</td>
<td>93 (96.9%)</td>
</tr>
<tr>
<td>NIHSS, mean (SD; range)</td>
<td>3.9 (3.0;0-12)</td>
</tr>
<tr>
<td>MoCA*, mean (SD; range)</td>
<td>26.0 (3.5;15-30)</td>
</tr>
</tbody>
</table>

NIHSS: National Institute of Health Stroke Scale (0-42); MoCA: Montreal Cognitive Assessment (0-30); *17 people were not included in the calculation of the mean MoCA score due to aphasia
Table 3.3 Distribution of REACH scale scores

<table>
<thead>
<tr>
<th>REACH Scale Levels</th>
<th>Dominant Side Affected</th>
<th>Non-dominant Side Affected</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td><strong>Mean Score (SD)</strong></td>
<td><strong>2.6 (1.4)</strong></td>
<td><strong>2.3 (1.6)</strong></td>
</tr>
</tbody>
</table>

SD: Standard Deviation
Table 3.4 Shapiro Wilk test of normality for the REACH scores

<table>
<thead>
<tr>
<th>Scale</th>
<th>Shapiro Wilk Statistic</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>REACH Scales Combined</td>
<td>0.912</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>REACH dominant affected scale</td>
<td>0.899</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>REACH non-dominant affected scale</td>
<td>0.922</td>
<td>0.006</td>
</tr>
</tbody>
</table>
Table 3.5 Interrater reliability and measurement error values

<table>
<thead>
<tr>
<th>Variables</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC (2,1)</td>
<td>0.968 (95% CI: 0.949-0.980, p&lt;0.001)</td>
</tr>
<tr>
<td>Linear Weighted Kappa ($\kappa_w$)</td>
<td>0.908 (95% CI: 0.893-0.926, p&lt;0.001)</td>
</tr>
<tr>
<td>SEM</td>
<td>0.27 (11.1% of mean)</td>
</tr>
<tr>
<td>MDC$_{95}$</td>
<td>0.75 (30.7% of mean)</td>
</tr>
</tbody>
</table>

ICC (2,1): intraclass correlation coefficient as captured in 2 way random effects model with absolute agreement; SEM: standard error of measurement; MDC$_{95}$: minimal detectable change calculated at a 95% confidence interval
Table 3.6 distribution of clinical measures used in validity analyses

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Mean (SD; range)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UE use measures</strong></td>
<td></td>
</tr>
<tr>
<td>MAL</td>
<td>1.9 (1.8; 0-5)</td>
</tr>
<tr>
<td>Affected UE Activity Counts</td>
<td>108 540 (80 325; 4353-324 593)</td>
</tr>
<tr>
<td><strong>UE function measures</strong></td>
<td></td>
</tr>
<tr>
<td>ARAT</td>
<td>29.1 (24.0; 0-57)</td>
</tr>
<tr>
<td>SIS-hand</td>
<td>38.1 (33.8; 0-100)</td>
</tr>
<tr>
<td><strong>UE impairment measures</strong></td>
<td></td>
</tr>
<tr>
<td>Chedoke-arm and hand</td>
<td>9.0 (4.5; 2-14)</td>
</tr>
<tr>
<td>Chedoke-shoulder pain</td>
<td>5.8 (1.5; 1-7)</td>
</tr>
</tbody>
</table>

SD: standard deviation; UE: upper extremity; MAL: Motor Activity Log; UE: upper extremity; ARAT: Action Research Arm Test; SIS-hand: Stroke Impact Scale-hand scale; Chedoke-arm and hand: Chedoke-McMaster arm and hand scales; Chedoke-shoulder pain: Chedoke-McMaster should pain scale
### Table 3.7 Correlation coefficients for validity analysis

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Spearman rank correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UE use measures</strong></td>
<td></td>
</tr>
<tr>
<td>MAL</td>
<td>$\rho=0.94$, $p&lt;0.001$</td>
</tr>
<tr>
<td>Affected UE Activity Counts</td>
<td>$\rho= 0.61$, $p&lt;0.001$</td>
</tr>
<tr>
<td><strong>UE function measures</strong></td>
<td></td>
</tr>
<tr>
<td>ARAT</td>
<td>$\rho=0.93$, $p&lt;0.001$</td>
</tr>
<tr>
<td>SIS-hand</td>
<td>$\rho=0.94$, $p&lt;0.001$</td>
</tr>
<tr>
<td><strong>UE impairment measures</strong></td>
<td></td>
</tr>
<tr>
<td>Chedoke-arm and hand</td>
<td>$\rho=0.91$, $p&lt;0.001$</td>
</tr>
<tr>
<td>Chedoke-shoulder pain*</td>
<td>$\rho=0.24$, $p=0.02$</td>
</tr>
</tbody>
</table>

UE: upper extremity; MAL: Motor Activity Log; UE: upper extremity; ARAT: Action Research Arm Test; SIS-hand: Stroke Impact Scale-hand scale; Chedoke-arm and hand: Chedoke-McMaster arm and hand scales; Chedoke-shoulder pain: Chedoke-McMaster should pain scale

* the correlation coefficients of the dominant scale was $\rho=0.11$, $p=0.42$, and for non-dominant scale was $\rho=0.29$, $p=0.06$
Figures

Figure 1.1 Sigmoidal relationship between UE use and UE function
Figure 1.2 COSMIN taxonomy of measurement properties

Adapted from Mokkink et al. (2010)
Figure 1.3 Conceptualization of validity for thesis

Validity

Construct

Hypothesis
Testing
Figure 2.1 Graphic representation of observed, important and detectable change

- **Observed Change**: Quantifies change that is observed from one time period to another or before and after a treatment. Traditionally expressed by an effect size (e.g., PES or SRM).
- **Detectable Change**: Captures the amount of change needed to be observed in order to surpass measurement error. Traditionally expressed as MDC_{90}, MDC_{95} or LOA value.
- **Important Change**: Captures the amount of observed change that is considered to be important. Traditionally expressed by a MCID value.

Note: MDC values can also be used to approximate important change (i.e., as a distribution-based method).

Distribution methods are statistically derived estimates (e.g., MDC, effect size). Anchor methods use external indicators (e.g., patient or clinician opinions or comparative measures).

MDC90: minimal detectable change (with 90% confidence interval); MDC95: minimal detectable change (with 95% confidence interval); LOA: limits of agreement; PES: population effect size; SRM: standardized response mean; MCID: minimally clinically important difference
Figure 2.2 Flow diagram of process to select final list of outcome measures

OM: outcome measure; AMAT: Arm Motor Activity Test; ARAT: Action Research Arm Test; CAHAI: Chedoke Arm and Hand Activity Inventory; Duruoz: Duruoz Hand Index; Frenchay: Frenchay Arm Test; FTHUE: Functional Test for the Hemiplegic Upper Extremity; Jebsen: Jebsen Hand Function Test; MAL: Motor Activity Log; SIS: Stroke Impact Scale; TEMPA: Upper Extremity Performance Scale for the Elderly; Wolf: Wolf Motor Function Test
Figure 2.3 Effect sizes by measure calculated at <3 months and ≥ 3 months post stroke

The bars on the graph represent the range of effect sizes calculated from studies that measured UE function across time. The full range of the effect sizes for the Frenchay is 0.2-5. Abbreviations: Wolf: Wolf Motor Function Test; Duruoz: Duruoz Hand Index; TEMPA: Upper Extremity Performance Scale for the Elderly; FTHUE: Functional Test for the Hemiplegic Upper Extremity; Jebsen: Jebsen Hand Function Test; CAHAI: Chedoke Arm and Hand Inventory; SIS: Stroke Impact Scale (hand scale); ARAT: Action Research Arm Test; Frenchay: Frenchay Arm Test
Figure 2.4 Comparison of observed change captured by ‘lab-based’ versus ‘patient-perceived’ functional measures

Points on the graph represent the effect sizes obtained from a single study. Lines on the graph represent a 1:1 relationship between the ‘lab-based’ vs ‘patient-perceived’ functional measures. Lab-based measures are located on the X-axes (ie. Wolf, ARAT). Patient-perceived functional measures are located on the Y axes (ie. MAL).

Abbreviations: MAL: Motor Activity Log; AOU: amount of use scale; QOM: quality of movement scale; CIMT: ARAT: Action Research Arm Test; Wolf: Wolf Motor Function Test; FAS: functional ability scale
Figure 2.5 Comparison of MDC\textsubscript{95} values relative to the sample means and scale maximums

The bars beside each measure represent the range of MDC\textsubscript{95} values extracted or calculated from different studies. *Two studies analysed different subsets of the same sample to obtain multiple estimates. ϿEstimates for the ARAT, Wolf (FAS) and SIS are missing from this graph as sample means were not provided in two studies. Abbreviations: MDC\textsubscript{95}: Minimally detectable change (with a 95% confidence interval) expressed as a percentage of A: the scale maximum score and B: the sample means; MAL: Motor Activity Log; AOU: amount of use scale; SIS: Stroke Impact Scale (hand); QOM: quality of movement scale; TEMPA: Upper Extremity Performance Scale for the Elderly; AMAT: Arm Motor Ability Test; FAS: functional.
Figure 3.1 Rating of Everyday Arm-use in the Community and Home (REACH) scale

<table>
<thead>
<tr>
<th>Level</th>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No use/Exercise only</td>
<td>□ Does not use the affected side at all or only uses the affected side for exercise purposes</td>
</tr>
</tbody>
</table>
| 1     | Stabilize only | □ Only uses the affected side to help stabilize objects during two-handed activities (in at least one example below)  
                      □ Stabilize item on surface (e.g., steady paper while writing)  
                      □ Uses affected hand as a passive vise or clamp (e.g., stabilize medication bottle in hand to enable unaffected side to open lid, stabilize material while buttoning or zipping)  
                      □ Stabilize item in between arm and body. |
| 2     | Assist or easy reach | □ Uses the affected side (actively) to assist the less affected side when performing two-handed activities, with the less affected side doing most of the work (in at least one example below).  
                      □ Assists stronger side when dressing and bathing (e.g., pulling up pants, unfold and orient clothes to put them on, assists in washing hair and/or body, getting clothes off hanger)  
                      □ Actively grasp items to allow stronger side to manipulate them (e.g., hold food item while stronger side chops, hold wallet while stronger side gets money out, hold jar while stronger side opens lid)  
                      □ Uses the weaker side to perform easy reach tasks that require minimal manipulation. (e.g., flips a light switch, closing cupboard, pushes door or pedestrian signal, turn on/off lever tap, especially if located on weaker side or stronger hand is holding something)  
                      □ Uses the affected side to perform some reach and grasp tasks which require active grasp or hand manipulation (in at least one example below).  
                      □ Reaching for light items (e.g., in fridge, cupboard/closet, drawer, or on table)  
                      □ Drinking from a glass or mug (can be cold drinks or half full drinks only)  
                      □ Cutting or chopping food (can be rough chopping or easy-to-cut foods only)  
                      □ Writing (daily) |
| 3     | Some reach and grasp with hand manipulation | □ Uses the affected side for all daily activities except those that involve heavy objects, high height, precision, fine motor or has negative consequences if performed inaccurately, (e.g., shaving, cleaning self after toilet, carrying something hot across a room, typing/using mouse)  
                      □ When reaching and grasping, the affected hand is used an equal or greater amount of time over a typical day compared to the non-affected hand. |
| 4     | Everyday use unless potential negative consequences | □ Uses the affected side for all tasks previously performed prior to the stroke (performance may be slower or less accurate than before the stroke)  
                      □ Uses the affected side for all tasks previously performed prior to the stroke (performance may be slower or less accurate than before the stroke)  
                      □ Does not use adaptive strategies for any tasks (i.e. compensates by using less affected side to avoid consequences or no longer performs a particular task because of fear of consequences) |
| 5     | Full Use | □ Uses the affected side for all tasks previously performed prior to the stroke (performance may be slower or less accurate than before the stroke)  
                      □ Does not use adaptive strategies for any tasks (i.e. compensates by using less affected side to avoid consequences or no longer performs a particular task because of fear of consequences) |
Rating of Everyday Arm-use in the Community and Home (REACH) Scale

Dominant side affected Algorithm

Do you use your R/L (affected) side to reach for things? (eg. flip a light switch, turn on/off lever tap)

Yes → Levels 2,3,4,5

No → Levels 0 or 1

Do you use your R/L (affected) side to help hold things in place?

No → Level 0, Verify with checklist

Yes → Level 1, Verify with checklist

Do you use your R/L (affected) side like you used it before the stroke?

No → Level 5, Verify with checklist

Yes → Levels 2,3,4

Do you use your R/L (affected) side to do some tasks you did before the stroke such as drink from a cup, pour a glass of juice or cut your food?

No → Level 2, Verify with checklist

Yes → Levels 3,4

Over a typical day, do you reach, grasp and manipulate objects with your R/L (affected) as much as you reach, grasp and manipulate objects with your R/L (less affected) side?

No → Level 3, Verify with checklist

Yes → Level 4, Verify with checklist
## Rating of Everyday Arm-use in the Community and Home (REACH) Scale

### Non-dominant side affected checklist

<table>
<thead>
<tr>
<th>Level</th>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>None/ Exercise only</td>
<td>□ Does not use the affected side at all or only uses the affected side for exercise purposes</td>
</tr>
<tr>
<td>1</td>
<td>Stabilize only</td>
<td>□ Only uses the affected side to help stabilize objects during two-handed activities <em>(in at least one example below)</em> &lt;br&gt; • Stabilize item on surface (e.g. steady paper while writing) &lt;br&gt; • Uses affected hand as passive vise or clamp (e.g. stabilize medication bottle in hand to enable stronger side to open lid, stabilize material while buttoning or zipping) &lt;br&gt; • Stabilize item in between arm and body.</td>
</tr>
<tr>
<td>2</td>
<td>Limited assist or easy reach</td>
<td>□ Uses the affected side (actively) during two handed activities that require minimal hand manipulation with the less affected side performing most of the work (e.g. assist to wash one side of body, assist to lift heavier items up, put shirt on, wash both hands under tap) &lt;br&gt; OR &lt;br&gt; □ Uses the weaker side to perform simple reach tasks that require minimal hand manipulation (e.g. flip a wall light switch, press walk sign or automatic door opener, close cupboard, turn on/off lever tap especially if located on affected side or less affected side is busy)</td>
</tr>
<tr>
<td>3</td>
<td>Some reach and grasp with hand manipulation</td>
<td>□ Uses the affected side to assist the less affected side with some activities that involve reaching, active grasp and/or hand manipulation <em>(in at least one example below)</em> &lt;br&gt; • Assists in carrying items with both hands (e.g. dish or pot, large grocery item) &lt;br&gt; • Assists stronger side when dressing and bathing (e.g. pull up pants, unfold and orient clothes to put them on, uses hand to assist in washing hair and/or body, get clothes off hanger) &lt;br&gt; • Actively grasp items to allow stronger side to manipulate them (e.g. hold food items while chopping, hold wallet while stronger side gets money out, hold jar while stronger side opens lid, do up seatbelt) &lt;br&gt; • Performs simple reach and grasp tasks when convenient and when there are no serious or inconvenient consequences if performed inaccurately (e.g. open a cupboard, door, drawer, fridge)</td>
</tr>
<tr>
<td>4</td>
<td>Everyday use unless potential negative consequences</td>
<td>□ Uses the affected side in a similar fashion as before the stroke in order to complete activities that traditionally require two hands or to finish tasks faster (e.g. both hands working together to complete one task or each hand is performing separate tasks) <em>(in all examples below)</em> &lt;br&gt; □ Uses the affected side in a similar fashion as before the stroke during dressing and bathing (e.g. contributes about the same amount as before the stroke when buttoning up clothes, putting on socks, tying shoelaces, drying self with towel) &lt;br&gt; □ Reaches for various items to help finish tasks faster (e.g. emptying dishwasher, putting groceries away, pick up things with both hands when tidying a room) &lt;br&gt; □ Reaches for various items if located on affected side (e.g. from a cupboard, drawer, table) &lt;br&gt; □ Uses affected side for support (e.g. pushing up from a chair, leaning on affected side if reaching for something far with less affected side, on the bus, use a handrail or grab bar) AND &lt;br&gt; □ Uses the affected side to assist in all daily activities <em>(as before the stroke)</em> except those that involve heavy objects, high height, precision, fine motor or has negative consequences if performed inaccurately (e.g. typing, carrying a hot drink across the room, unlocking door with key)</td>
</tr>
<tr>
<td>5</td>
<td>Full Use</td>
<td>□ Uses the affected side for all tasks previously performed prior to the stroke (performance may be slower or less accurate than before the stroke) AND □ Does not use adaptive strategies for tasks (compensates by using less affected side to avoid consequences or no longer performs a particular task because of fear of consequences)</td>
</tr>
</tbody>
</table>
Rating of Everyday Arm-use in the Community and Home (REACH) Scale

Non-dominant side affected algorithm

Do you use your R/L (affected) side during two handed activities (i.e., to help the unaffected side) such as washing your hair, getting dressed, lifting items up?

No → Levels 0 or 1

Do you use your R/L (affected) side to help hold things in place?

No → Level 0
Verify with checklist

Yes → Level 1
Verify with checklist

Yes, Levels 2, 3, 4, 5

Do you use your R/L (affected) side like you used it before the stroke?

No → Levels 2, 3, 4

Yes, Level 5
Verify with checklist

If getting something out of a cupboard/drawer, do you ever use one side to open the cupboard/drawer and the other side to reach for the item?

cue: what if cupboard/drawer is on R/L (affected) side?

No → Level 2
Verify with checklist

Yes → Levels 3, 4

When bathing and dressing, do you use your R/L (affected) side in a similar fashion as before the stroke?

Note: movements with R/L (affected) side can be slower and less accurate than before the stroke

No → Level 3
Verify with checklist

Yes → Level 4
Verify with checklist
Figure 3.2 Plot of rater agreement

Note: Plots include a horizontal jitter to ensure visibility of data points at identical locations.
Figure 3.3 Scatterplot of REACH scale scores with external measures of UE use, UE function and UE impairment

Note: Plots include a horizontal jitter to ensure visibility of data points at identical locations.

- Non-dominant
- Dominant
References


Kain ZN, MacLaren J. P less than. 05: What does it really mean? Pediatrics. 2007;119:608-610.


Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychol Methods. 2002;7:105-125.


Rehme AK, Fink GR, von Cramon DY, Grefkes C. The role of the contralesional motor cortex for motor recovery in the early days after stroke assessed with longitudinal FMRI. Cereb Cortex. 2011;21:756-768.


Wu CY, Chuang LL, Lin KC, Chen HC, Tsay PK. Randomized trial of distributed constraint induced therapy versus bilateral arm training for the rehabilitation of upper-limb

## Appendix A: Factors associated with arm use

<table>
<thead>
<tr>
<th>Factor</th>
<th>Outcome Measures</th>
<th>Study Design</th>
<th>Relationship</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UE Impairments</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>UE AROM</strong></td>
<td><em>AROM</em>: kinematic analysis&lt;br&gt;Use: Accelerometers-duration of use of affected and non-affected UE</td>
<td>N=34&lt;br&gt;&lt; 1 month post stroke&lt;br&gt;Data Analysis: Spearman rank correlation</td>
<td>Elbow and wrist AROM were significantly associated with use (ρ=0.50-0.63, p&lt;0.01)</td>
<td>Lang et al. 2007</td>
</tr>
<tr>
<td><strong>UE Strength</strong></td>
<td><em>Strength</em>: grip and composite of hand-held dynamometry of shoulder, wrist, elbow&lt;br&gt;<strong>UE use</strong>: MAL</td>
<td>N=93&lt;br&gt;&gt;6 months post stroke&lt;br&gt;Data Analysis: Spearman rank correlation and Stepwise Multiple Regression with 5 impairment factors (composite strength, grip strength, pain, spasticity)</td>
<td>Composite strength was only factor associated with arm use in the model (R²=0.78, p=0.03). Grip strength was not a significant factor in the model but was significantly correlated to UE use (r=0.61, p&lt;0.01)</td>
<td>Harris and Eng 2007</td>
</tr>
<tr>
<td></td>
<td><em>Strength</em>: grip; hand-held dynamometry of shoulder, elbow and wrist&lt;br&gt;<strong>Use</strong>: Accelerometers-duration of use of affected and unaffected UE</td>
<td>N=34&lt;br&gt;&lt;1 month post stroke&lt;br&gt;Data Analysis: Spearman rank correlation</td>
<td>Elbow, wrist and grip strength were significantly correlated with use (ρ=0.37-0.52, p&lt;0.05)</td>
<td>Lang et al. 2007</td>
</tr>
<tr>
<td></td>
<td><em>Strength</em>: MI&lt;br&gt;<strong>Use</strong>: Accelerometers: activity counts of affected UE</td>
<td>N=38&lt;br&gt;1 day, 1 week, 3 months, 6 months post stroke&lt;br&gt;Data Analysis: Pearson correlation</td>
<td>Strength was correlated with use at 1 day (r=0.44, p=0.01) and 1 week (r=0.5, p=0.01) and not significantly correlated at 3 and 6 months</td>
<td>Reiterer et al. 2008</td>
</tr>
<tr>
<td>Factor</td>
<td>Outcome Measures</td>
<td>Study Design</td>
<td>Relationship</td>
<td>Reference</td>
</tr>
<tr>
<td>--------</td>
<td>------------------</td>
<td>--------------</td>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Composite UE impairment (ie. Fugl Meyer)</td>
<td><strong>UE impairment:</strong> FM Use: Accelerometers-affected UE activity counts; ratio of affected UE activity/unaffected UE activity</td>
<td>N=39 &lt;1 month post stroke Data Analysis: Spearman rank correlation</td>
<td>UE impairment was correlated to UE use ($\rho=0.69$, affected UE use; $\rho=0.54$, UE use ratio, p&lt;0.001)</td>
<td>Gebruers et al. 2008</td>
</tr>
<tr>
<td>Composite UE impairment (ie. Fugl Meyer)</td>
<td><strong>UE impairment:</strong> FM Use: MAL</td>
<td>N=57 &gt;6 months post stroke Data Analysis: Stepwise multiple linear regression with 5 predictors of use following CIMT inputted into model (age, time post stroke, stroke severity, UE impairment, spasticity.</td>
<td>UE impairment was a significant predictor in the model ($\beta=0.013$, p&lt;0.02). UE impairment and age explained 20% of variance in amount of arm use. UE impairment and time post stroke explained 43% of variance in quality of arm use</td>
<td>Lin et al. 2009d</td>
</tr>
<tr>
<td></td>
<td><strong>UE impairment:</strong> FM Use: Accelerometers-ratio of affected/unaffected arm use</td>
<td>N=27 &gt;6 months post stroke Data Analysis: Spearman rank correlation; comparison of logarithmic and linear regression</td>
<td>UE impairment was significantly correlated to ratio of arm use ($\rho=0.75$, p&lt;0.001). Logarithmic model of UE impairment explained more variance in arm use than linear model. $R^2=0.72$ vs $R^2=0.58$, p&lt;0.001</td>
<td>Michielson et al. 2009</td>
</tr>
<tr>
<td></td>
<td><strong>UE impairment:</strong> FM Use: MAL</td>
<td>N=30 &gt;6 months post stroke Data Analysis: MANOVA</td>
<td>Increase in arm use was different among mild and severely affected groups (improvement of 1.0 vs improvement of 0.2, p&lt;0.065)</td>
<td>Pang et al. 2006</td>
</tr>
<tr>
<td>Factor</td>
<td>Outcome Measures</td>
<td>Study Design</td>
<td>Relationship</td>
<td>Reference</td>
</tr>
<tr>
<td>---------</td>
<td>---------------------------------------------------------------------------------</td>
<td>-----------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>UE Pain</td>
<td><strong>UE impairment</strong>: FM Use: Accelerometers-duration of use ratio (ratio of nonaffected UE/affected UE, duration of affected UE use)</td>
<td>N=31 &lt;1 month post stroke Data Analysis: Spearman rank correlation</td>
<td>UE impairment was significantly correlated to duration of affected UE arm use ($\rho=0.60$, $p&lt;0.001$) and UE use ratio ($\rho=-0.85$, $p&lt;0.001$)</td>
<td>Thrane et al. 2011</td>
</tr>
<tr>
<td>Spasticity</td>
<td><strong>Spasticity</strong>: MAS Use: MAL</td>
<td>N=93 &gt;6 months post stroke Data Analysis: Stepwise Multiple Regression with 5 impairment factors (composite strength, grip strength, pain, spasticity); Spearman rank correlation</td>
<td>Spasticity was not a significant factor in the model. Spasticity was significantly correlated to UE use ($\rho=-0.71$, $p&lt;0.01$)</td>
<td>Harris and Eng 2007</td>
</tr>
<tr>
<td>UE Pain</td>
<td><strong>Pain</strong>: BPI Use: MAL</td>
<td>N=93 &gt;6 months post stroke Data Analysis: Stepwise Multiple Regression with 5 impairment factors (composite strength, grip strength, pain, spasticity); Spearman rank correlation</td>
<td>Pain was not a significant factor in the model nor was significantly correlated to UE use ($r=-0.06$, $p&gt;0.05$)</td>
<td>Harris and Eng 2007</td>
</tr>
<tr>
<td></td>
<td><strong>Pain</strong>: Shoulder pain on 0-10 scale Use: Accelerometers-duration of use of affected and nonaffected UE</td>
<td>N=34 &lt;1 month post stroke Spearman rank correlation</td>
<td>Affected shoulder pain was significantly correlated to UE use ($\rho=0.41$, $P&lt;0.05$)</td>
<td>Lang et al. 2007</td>
</tr>
<tr>
<td>Factor</td>
<td>Outcome Measures</td>
<td>Study Design</td>
<td>Relationship</td>
<td>Reference</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------------------</td>
</tr>
</tbody>
</table>
| Stroke severity      | *Stroke severity: NIHSS*  
*Use: Accelerometers-*affected UE activity counts; ratio of affected UE activity/unaffected UE activity | N=39  
<1 month post stroke  
Data Analysis: Spearman rank correlation | Stroke severity was correlated to UE use  
\( \rho = -0.75 \) (affected UE use)  
\( \rho = -0.59 \) (UE use ratio)  
p<0.001 | Gebruers et al. 2008 |
|                      | *Stroke severity: NIHSS*  
*Use: MAL*                                                                 | N=57  
>6 months post stroke  
Data Analysis: Stepwise multiple linear regression with 5 predictors of use following CIMT inputted into model (Age, time post stroke, stroke severity, UE impairment, spasticity.) | Stroke severity was not a significant predictor of amount or quality of arm use following CIMT | Lin et al. 2009d   |
|                      | *Stroke severity: NIHSS*  
*Use: Accelerometers-*activity counts of affected UE | N=38  
Measured 1 day, 1 week, 3 months, 6 months post stroke  
Data Analysis: Spearman rank correlation | Stroke severity was not significantly correlated to use at any time points | Reiterer et al. 2008 |
| UE Sensation         | *Sensation: light touch*  
*UE use: MAL*                                                                 | N=93  
>6 months post stroke  
Data Analysis: Stepwise Multiple Regression with 5 impairment factors (composite strength, grip strength, pain, spasticity);  
Spearman rank correlation | Sensation was not a significant factor in the model. Sensation was significantly correlated to UE use (\( \rho = -0.43 \), p<0.01) | Harris and Eng 2007 |
<table>
<thead>
<tr>
<th>Factor</th>
<th>Outcome Measures</th>
<th>Study Design</th>
<th>Relationship</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensation (light touch)</td>
<td>Use (Accelerometers-duration of affected and non-affected UE)</td>
<td>N=34 &lt;1 month post stroke</td>
<td>Sensation was not significantly correlated to UE use ($\rho=-0.15, p&lt;0.01$)</td>
<td>Lang et al. 2007</td>
</tr>
<tr>
<td><strong>UE Function</strong></td>
<td><strong>Performance-based (e.g. ARAT, Wolf)</strong></td>
<td>Function: ARAT Use: MAL</td>
<td>Pre-treatment function was correlated to post treatment use ($\rho=0.62-0.66, p&lt;0.001$)</td>
<td>Chen et al. 2012</td>
</tr>
<tr>
<td></td>
<td>Function: ARAT Use: MAL</td>
<td>N=39 &lt; 6 months post stroke</td>
<td>Function was correlated with use ($r=0.61, p$ value not provided)</td>
<td>Dromerick et al. 2006</td>
</tr>
<tr>
<td></td>
<td>Function: ARAT, Wolf Use: Accelerometers-duration of use of affected UE</td>
<td>N=34 &lt; 1 month post stroke</td>
<td>Both measures of function were significantly correlated with use ($\rho=0.40-0.65, P&lt;0.05$)</td>
<td>Lang et al. 2007</td>
</tr>
<tr>
<td></td>
<td>Function: ARAT Use: Accelerometers-ratio of affected/unaffected UE use</td>
<td>N=27 &gt;6 months post stroke</td>
<td>Function was significantly correlated with use ($\rho=0.71, p&lt;0.001$). Logarithmic model of function explained more variance in arm use than linear model ($R^2=0.66$ vs $R^2=0.57$, difference wasn’t significant)</td>
<td>Michielson et al. 2009</td>
</tr>
<tr>
<td>Factor</td>
<td>Outcome Measures</td>
<td>Study Design</td>
<td>Relationship</td>
<td>Reference</td>
</tr>
<tr>
<td>----------------------------</td>
<td>------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Performance-based (e.g. ARAT, Wolf)</td>
<td>Function: ARAT Use: MAL</td>
<td>N=56 &gt;6 months post stroke Data Analysis: Spearman rank correlation</td>
<td>Function was significantly correlated to use ($\rho=0.63$, $p&lt;0.001$)</td>
<td>Van der Lee et al. 2004</td>
</tr>
<tr>
<td></td>
<td>Function: CAHAI Use: MAL</td>
<td>N=93 &gt; 6 months post stroke Data Analysis: Pearson’s correlation</td>
<td>Function was significantly correlated to use ($r=0.82$, $p&lt;0.01$)</td>
<td>Harris et al. 2007</td>
</tr>
<tr>
<td></td>
<td>Function: Wolf-time Use: MAL</td>
<td>N=48 &lt;6 months post stroke Data Analysis: Bayesian regression to compare fit of different models</td>
<td>Non-linear sigmoidal model fit relationship b/w UE function and UE use better than linear model</td>
<td>Hidaka et al. 2012</td>
</tr>
<tr>
<td></td>
<td>Function: Wolf Use: MAL</td>
<td>N=169 &lt;6 months post stroke Data Analysis: Computer simulations and EXCITE trial data to test hypothesis that a functional threshold for increase in UE use following therapy exists; Multiple stepwise linear regressions to see if function right after therapy stayed in model</td>
<td>Wolf FAS score at 1 week post treatment and wolf strength task were only 2 predictors of change in MAL over time ($R^2=1.123$, $p&lt;0.005$). Simulation model provides early support that a functional threshold exists.</td>
<td>Schweighofer et al. 2009</td>
</tr>
<tr>
<td>Factor</td>
<td>Outcome Measures</td>
<td>Study Design</td>
<td>Relationship</td>
<td>Reference</td>
</tr>
<tr>
<td>--------</td>
<td>------------------</td>
<td>--------------</td>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td><strong>Self-report (e.g. SIS, ABILHAND)</strong></td>
<td>Function: ABILHAND  &lt;br&gt; Use: Accelerometers-ratio of affected arm use/unaffected arm use</td>
<td>N=27  &lt;br&gt; &gt;6 months post stroke  &lt;br&gt; Data Analysis: Spearman rank correlation; comparison of logarithmic and linear regression models</td>
<td>Function was significantly correlated with use (ρ=0.64, p&lt;0.01). Logarithmic model and linear model explained similar amount of variance in arm use ratio (R²=0.33 vs R²=0.30)</td>
<td>Michielson et al. 2009</td>
</tr>
<tr>
<td><strong>Self-report (e.g. SIS, ABILHAND)</strong></td>
<td>Function: SIS-hand  &lt;br&gt; Use: MAL</td>
<td>N=222  &lt;br&gt; &lt;6 months post stroke  &lt;br&gt; Data Analysis: Pearson correlation coefficient</td>
<td>Function was significantly correlated with use (r=0.68-0.72, p&lt;0.01)</td>
<td>Uswatte et al. 2006a</td>
</tr>
<tr>
<td></td>
<td>Function: SIS-hand  &lt;br&gt; Use: Accelerometers-activity counts of affected UE; ratio of affected/nonaffected activity counts</td>
<td>N=45  &lt;br&gt; &gt;6 months post stroke  &lt;br&gt; Data Analysis: Spearman rank correlation</td>
<td>Function was significantly correlated with ratio of use (ρ=0.58, p&lt;0.01) and affected UE use (ρ=0.61, p&lt;0.01).</td>
<td>Van der Pas et al. 2011</td>
</tr>
<tr>
<td><strong>Personal Factors</strong></td>
<td><strong>Use</strong> : MAL</td>
<td>N=39  &lt;br&gt; &lt;6 months post stroke  &lt;br&gt; Data Analysis: t-tests</td>
<td>No significant differences between groups (ie dominant side affected vs non-dominant side affected)</td>
<td>Dromerick et al. 2006</td>
</tr>
<tr>
<td>Concordance of dominant side and side of stroke</td>
<td><strong>Use</strong>: MAL</td>
<td>N=55  &lt;br&gt; &gt;6 months post stroke  &lt;br&gt; Data Analysis: Stepwise Linear regression with 6 personal factors</td>
<td>Concordance of dominant side and side of stroke not a significant predictor of use following CIMT</td>
<td>Fritz et al. 2006</td>
</tr>
<tr>
<td>Factor</td>
<td>Outcome Measures</td>
<td>Study Design</td>
<td>Relationship</td>
<td>Reference</td>
</tr>
<tr>
<td>-------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Use: Accelerometers</td>
<td>duration of use of affected UE only; unaffected UE only; bilateral UE</td>
<td>N=60</td>
<td>Arm use patterns were significantly different between groups (F_{2,112}=9.6430, P=.0001). Unaffected UE use and bilateral UE use greater in non-dominant group vs dominant group (P=0.002); P=0.023); Affected UE use greater in dominant group vs non-dominant group (P=0.001)</td>
<td>Haaland et al. 2012</td>
</tr>
<tr>
<td>Concordance of dominant side and side of stroke</td>
<td>Use: Accelerometers-duration of unaffected, affected and bilateral UE use while performing AMAT test</td>
<td>N=29</td>
<td>Patterns of use differed significantly based on side of stroke and side of hand dominance (F(1,27)=5.29, p&lt;0.05). Non-dominant hand affected group used their non-affected UE more than non-dominant hand group. Also dominant hand affected group used their hands bilaterally more often than non-dominant hand affected group. Groups did not differ in duration of affected arm use</td>
<td>Rinehart et al. 2009</td>
</tr>
<tr>
<td>Factor</td>
<td>Outcome Measures</td>
<td>Study Design</td>
<td>Relationship</td>
<td>Reference</td>
</tr>
<tr>
<td>--------</td>
<td>------------------</td>
<td>--------------</td>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Use: BART</td>
<td>N=24 &gt;6 months post stroke Data Analysis: Frequency analysis</td>
<td>11/13 participants with dominant UE showed non-use vs 2/11 participants with non-dominant side affected</td>
<td>Han et al. 2013</td>
<td></td>
</tr>
<tr>
<td>Use: MAL</td>
<td>N=169 &gt;6 months post stroke Data Analysis: Stepwise multiple linear regression</td>
<td>Concordance was not a significant factor in model</td>
<td>Schweighofer et al. 2009</td>
<td></td>
</tr>
<tr>
<td>Use: Accelerometers-duration of affected UE use/duration of non-affected UE use</td>
<td>N=222 &lt; 6 months post stroke Data Analysis: Pearson correlation</td>
<td>People with dominant UE affected had greater correlation b/w self-report and objective measures (r=0.56-0.59) than did people with non-dominant UE affected (0.28-0.34)</td>
<td>Uswatte et al. 2006a</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Use: MAL</td>
<td>N=55 &gt;6 months post stroke Data Analysis: Stepwise multiple linear regression with 6 personal factors</td>
<td>Age was only significant predictor of use following CIMT</td>
<td>Fritz et al. 2006</td>
</tr>
<tr>
<td></td>
<td>Use: MAL</td>
<td>N=57 &gt;6 months post stroke Data Analysis: Stepwise multiple linear regression with 5 predictors in model (age, time post stroke, stroke severity, UE impairment, spasticity)</td>
<td>Age was a significant predictor of amount of use following CIMT in model but not quality of use.</td>
<td>Lin et al. 2009d</td>
</tr>
<tr>
<td>Factor</td>
<td>Outcome Measures</td>
<td>Study Design</td>
<td>Relationship</td>
<td>Reference</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td><strong>UE Impairments</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>UE AROM</strong></td>
<td>AROM: kinematic analysis &lt;br&gt;Use: Accelerometers-duration of use of affected and non-affected UE</td>
<td>N=34 &lt; 1 month post stroke &lt;br&gt;Data Analysis: Spearman rank correlation</td>
<td>Elbow and wrist AROM were significantly associated with use ($\rho=0.50-0.63$, p&lt;0.01)</td>
<td>Lang et al. 2007</td>
</tr>
<tr>
<td><strong>UE Strength</strong></td>
<td>Strength: grip and composite of hand-held dynamometry of shoulder, wrist, elbow &lt;br&gt;UE use: MAL</td>
<td>N=93 &gt;6 months post stroke &lt;br&gt;Data Analysis: Spearman rank correlation and Stepwise Multiple Regression with 5 impairment factors (composite strength, grip strength, pain, spasticity)</td>
<td>Composite strength was only factor associated with arm use in the model ($R^2=0.78$, p=0.03). Grip strength was not a significant factor in the model but was significantly correlated to UE use ($r=0.61$, p&lt;0.01)</td>
<td>Harris and Eng 2007</td>
</tr>
<tr>
<td></td>
<td>Strength: grip; hand-held dynamometry of shoulder, elbow and wrist &lt;br&gt;Use: Accelerometers-duration of use of affected and unaffected UE</td>
<td>N=34 &lt;1 month post stroke &lt;br&gt;Data Analysis: Spearman rank correlation</td>
<td>Elbow, wrist and grip strength were significantly correlated with use ($\rho=0.37-0.52$, p&lt;0.05)</td>
<td>Lang et al. 2007</td>
</tr>
<tr>
<td><strong>Strength</strong></td>
<td>MI &lt;br&gt;Use: Accelerometers: activity counts of affected UE</td>
<td></td>
<td>Strength was correlated with use at 1 day ($r=0.44$, p=0.01) and 1 week ($r=0.5$, p=0.01) and not significantly correlated at 3 and 6 months</td>
<td>Reiterer et al. 2008</td>
</tr>
<tr>
<td>Factor</td>
<td>Outcome Measures</td>
<td>Study Design</td>
<td>Relationship</td>
<td>Reference</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------</td>
<td>--------------</td>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Composite UE impairment (ie. Fugl Meyer)</td>
<td>Composite UE impairment: FM Use: Accelerometers-affected UE activity counts; ratio of affected UE activity/unaffected UE activity</td>
<td>N=39 &lt;1 month post stroke Data Analysis: Spearman rank correlation</td>
<td>UE impairment was correlated to UE use ($\rho=0.69$, affected UE use; $\rho=0.54$, UE use ratio, $p&lt;0.001$)</td>
<td>Gebruers et al. 2008</td>
</tr>
<tr>
<td>Composite UE impairment (ie. Fugl Meyer)</td>
<td>Composite UE impairment: FM Use: MAL</td>
<td>N=57 &gt;6 months post stroke Data Analysis: Stepwise multiple linear regression with 5 predictors of use following CIMT inputted into model (age, time post stroke, stroke severity, UE impairment, spasticity).</td>
<td>UE impairment was a significant predictor in the model ($\beta=0.013$, $p&lt;0.02$). UE impairment and age explained 20% of variance in amount of arm use. UE impairment and time post stroke explained 43% of variance in quality of arm use</td>
<td>Lin et al. 2009d</td>
</tr>
<tr>
<td>UE impairment: FM Use: Accelerometers-ratio of affected/unaffected arm use</td>
<td>N=27 &gt; 6 months post stroke Data Analysis: Spearman rank correlation; comparison of logarithmic and linear regression</td>
<td>UE impairment was significantly correlated to ratio of arm use ($\rho=0.75$, $p&lt;0.001$). Logarithmic model of UE impairment explained more variance in arm use than linear model. $R^2=0.72$ vs $R^2=0.58$, $p&lt;0.001$</td>
<td>Michielson et al. 2009</td>
<td></td>
</tr>
<tr>
<td>UE impairment: FM Use: MAL</td>
<td>N=30 &gt; 6 months post stroke Data Analysis: MANCOVA</td>
<td>Increase in arm use was different among mild and severely affected groups (improvement of 1.0 vs improvement of 0.2, $p&lt;0.065$)</td>
<td>Pang et al. 2006</td>
<td></td>
</tr>
<tr>
<td>Factor</td>
<td>Outcome Measures</td>
<td>Study Design</td>
<td>Relationship</td>
<td>Reference</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------------------------------------------------</td>
<td>-----------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>UE Pain</td>
<td><strong>UE impairment</strong>: FM</td>
<td>N=31</td>
<td>UE impairment was significantly correlated to duration of affected UE arm use ((\rho=0.60, p&lt;0.001)) and UE use ratio ((\rho=-0.85, p&lt;0.001))</td>
<td>Thrane et al. 2011</td>
</tr>
<tr>
<td></td>
<td><strong>Use</strong>: Accelerometers-duration of use ratio (ratio of nonaffected UE/affected UE, duration of affected UE use)</td>
<td>&lt;1 month post stroke Data Analysis: Spearman rank correlation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Spasticity</strong>: MAS</td>
<td>N=93</td>
<td>Spasticity was not a significant factor in the model. Spasticity was significantly correlated to UE use ((\rho=-0.71, p&lt;0.01))</td>
<td>Harris and Eng 2007</td>
</tr>
<tr>
<td></td>
<td><strong>Use</strong>: MAL</td>
<td>&gt;6 months post stroke Data Analysis: Stepwise Multiple Regression with 5 impairment factors (composite strength, grip strength, pain, spasticity); Spearman rank correlation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Pain</strong>: BPI</td>
<td>N=93</td>
<td>Pain was not a significant factor in the model nor was significantly correlated to UE use ((r=-0.06, p&gt;0.05))</td>
<td>Harris and Eng 2007</td>
</tr>
<tr>
<td></td>
<td><strong>Use</strong>: MAL</td>
<td>&gt;6 months post stroke Data Analysis: Stepwise Multiple Regression with 5 impairment factors (composite strength, grip strength, pain, spasticity); Spearman rank correlation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Pain</strong>: Shoulder pain on 0-10 scale</td>
<td>N=34</td>
<td>Affected shoulder pain was significantly correlated to UE use ((\rho=0.41, P&lt;0.05))</td>
<td>Lang et al. 2007</td>
</tr>
<tr>
<td></td>
<td><strong>Use</strong>: Accelerometers-duration of use of affected and nonaffected UE</td>
<td>&lt;1 month post stroke Spearman rank correlation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor</td>
<td>Outcome Measures</td>
<td>Study Design</td>
<td>Relationship</td>
<td>Reference</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Stroke severity</td>
<td>Stroke severity: NIHSS Use: Accelerometers-affected UE activity counts; ratio of affected UE activity/unaffected UE activity</td>
<td>N=39 &lt;1 month post stroke Data Analysis: Spearman rank correlation</td>
<td>Stroke severity was correlated to UE use $\rho=-0.75$ (affected UE use) $\rho=-0.59$ (UE use ratio) p&lt;0.001</td>
<td>Gebruers et al. 2008</td>
</tr>
<tr>
<td>Stroke severity</td>
<td>Stroke severity: NIHSS Use: MAL</td>
<td>N=57 &gt;6 months post stroke Data Analysis: Stepwise multiple linear regression with 5 predictors of use following CIMT inputted into model (Age, time post stroke, stroke severity, UE impairment, spasticity.</td>
<td>Stroke severity was not a significant predictor of amount or quality of arm use following CIMT</td>
<td>Lin et al. 2009d</td>
</tr>
<tr>
<td>Stroke severity</td>
<td>Stroke severity: NIHSS Use: Accelerometers-activity counts of affected UE</td>
<td>N=38 Measured 1 day, 1 week, 3 months, 6 months post stroke Data Analysis: Spearman rank correlation</td>
<td>Stroke severity was not significantly correlated to use at any time points</td>
<td>Reiterer et al. 2008</td>
</tr>
</tbody>
</table>
Appendix B: References of the included articles in chapter 2

The following 68 articles were included in the Systematic Review of the Responsiveness of UE functional measures following stroke: A focus on important, detectable and observed change:

Appendix C: Effect sizes capturing observed change over natural recovery

<table>
<thead>
<tr>
<th>Measure</th>
<th>Effect Size Statistic</th>
<th>Population Effect Size</th>
<th>Standardized Response Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N=# of studies; n=total # participants)</td>
<td>Baseline test &lt;1 month</td>
<td>Baseline test &gt;1 and &lt; 3 months</td>
<td>Baseline test ≥3 months</td>
</tr>
<tr>
<td>ABILHAND (N=1; n=45)</td>
<td>0.77-0.86</td>
<td></td>
<td>0.10</td>
</tr>
<tr>
<td>ARAT (N=14; n=615)</td>
<td>0.22-2.4</td>
<td>0.17-1.56</td>
<td>0.04-0.51</td>
</tr>
<tr>
<td>CAHAI (N=3; n=129)</td>
<td>0.66-0.90</td>
<td></td>
<td>0.04-0.27</td>
</tr>
<tr>
<td>Duruoz (N=1; n=56)</td>
<td>0.3-0.5</td>
<td>0.20-5.5</td>
<td>0.27</td>
</tr>
<tr>
<td>Frenchay (N=4; n=146)</td>
<td>0.2-0.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FTHUE (N=1; n=18)</td>
<td>0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jebsen (N=1; n=22)</td>
<td>2.4</td>
<td>0.44-0.73</td>
<td>0.09</td>
</tr>
<tr>
<td>SIS-hand (N=3; n=174)</td>
<td>0.25-1.02</td>
<td>0.01-0.14</td>
<td></td>
</tr>
<tr>
<td>TEMPA (N=1; n=132)</td>
<td>0.4</td>
<td></td>
<td>0.76 (&gt;1 and &lt; 3 months)</td>
</tr>
<tr>
<td>Wolf (N=1; n=53)</td>
<td>0.44</td>
<td>0.14-0.22</td>
<td>0.08</td>
</tr>
</tbody>
</table>

PES: population effect size; SRM: standardized response mean; ARAT: Action Research Arm Test; CAHAI: Chedoke Arm and Hand Activity Inventory; Duruoz: Duruoz Hand Index; Frenchay: Frenchay Arm Test; FTHUE: Functional Test for the Hemiplegic Upper Extremity; HFS: Hand Function Survey; Jebsen: Jebsen Hand Function Test; SIS: Stroke Impact Scale; Wolf: Wolf Motor Function Test; FAS: functional ability scale
## Appendix D: Minimal detectable change (MDC) and limits of agreement values by measure

<table>
<thead>
<tr>
<th>Measure</th>
<th>SEM</th>
<th>MDC&lt;sub&gt;90&lt;/sub&gt;</th>
<th>MDC&lt;sub&gt;95&lt;/sub&gt;</th>
<th>LOA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Score (%&lt;sub&gt;max&lt;/sub&gt;-%&lt;sub&gt;mean&lt;/sub&gt;)</td>
<td>Raw Score (%&lt;sub&gt;max&lt;/sub&gt;-%&lt;sub&gt;mean&lt;/sub&gt;)</td>
<td>Raw score (%&lt;sub&gt;max&lt;/sub&gt;-%&lt;sub&gt;mean&lt;/sub&gt;)</td>
<td>Raw score (%&lt;sub&gt;max&lt;/sub&gt;-%&lt;sub&gt;mean&lt;/sub&gt;)</td>
<td></td>
</tr>
<tr>
<td>Accelerometry</td>
<td>0.05 (5%; 8.8%)</td>
<td>0.12 (12%; 21.1%)</td>
<td>0.14 (14%; 24.6%)</td>
<td></td>
</tr>
<tr>
<td>AMAT Time</td>
<td>0.51 (0.43%; 15.1%)</td>
<td>1.2 (1.0%; 35.1%)</td>
<td>1.4 (1.2%; 41.8%)</td>
<td></td>
</tr>
<tr>
<td>FAS</td>
<td>0.19 (3.8%; 7.8%)</td>
<td>0.44 (8.8%; 18.4%)</td>
<td>0.53 (10.6%; 23.1%)</td>
<td></td>
</tr>
<tr>
<td>QOM</td>
<td>0.16 (2.3%; 6.9%)</td>
<td>0.37 (7.4%; 16.3%)</td>
<td>0.44 (8.8%; 19.4%)</td>
<td></td>
</tr>
<tr>
<td>ARAT&lt;sup&gt;11,76&lt;/sup&gt;</td>
<td>1.3 (2.2%; NC)</td>
<td>3.0 (5.3%; NC)</td>
<td>3.5 (6.1%; NC)</td>
<td>(-) 5.7 (+) 6.2</td>
</tr>
<tr>
<td>CAHAI-13&lt;sup&gt;13&lt;/sup&gt;</td>
<td>2.8 (3.1%; 5.9%)</td>
<td>6.3 (6.9%; 13.4%)</td>
<td>7.8 (8.6%; 16.5%)</td>
<td></td>
</tr>
<tr>
<td>CAHAI-9&lt;sup&gt;13&lt;/sup&gt;</td>
<td>2.6 (4.1%; 7.6%)</td>
<td>6.0 (9.5%; 17.6%)</td>
<td>7.1 (11.3%; 20.9%)</td>
<td></td>
</tr>
<tr>
<td>CAHAI-8&lt;sup&gt;13&lt;/sup&gt;</td>
<td>2.3 (4.1%; 7.5%)</td>
<td>5.3 (9.5%; 17.3%)</td>
<td>6.3 (11.3%; 20.5%)</td>
<td></td>
</tr>
<tr>
<td>CAHAI-7&lt;sup&gt;13&lt;/sup&gt;</td>
<td>2.3 (4.7%; 8.6%)</td>
<td>5.4 (11.0%; 17.6%)</td>
<td>6.4 (13.1%; 24.0%)</td>
<td></td>
</tr>
<tr>
<td>Duruoz&lt;sup&gt;26&lt;/sup&gt;</td>
<td>2.66 (3.0%; 8.5%)</td>
<td>6.19 (6.9%; 19.8%)</td>
<td>7.4 (8.2%; 23.6%)</td>
<td></td>
</tr>
<tr>
<td>MAL-28&lt;sup&gt;71&lt;/sup&gt;</td>
<td>0.46 (9.2%; 30.5%)</td>
<td>1.06 (21.2%; 70.7%)</td>
<td>1.27 (25.4%; 84.7%)</td>
<td></td>
</tr>
<tr>
<td>AOU</td>
<td>QOM</td>
<td>0.38 (7.6%; 25.5%)</td>
<td>0.89 (17.8%; 59.3%)</td>
<td>1.06 (21.2%; 70.7%)</td>
</tr>
<tr>
<td>MAL-14&lt;sup&gt;10,75&lt;/sup&gt;</td>
<td>0.37 (7.4%; 30.8%)</td>
<td>0.87 (17.4%; 72.5%)</td>
<td>1.04 (20.8%; 86.7%)</td>
<td>(-) 0.70-0.88; (+) 0.85 1.1 (17-22.0%)</td>
</tr>
<tr>
<td>AOU</td>
<td>QOM</td>
<td>0.24 (4.8%; 26.7%)</td>
<td>0.56 (11.2%; 62.2%)</td>
<td>0.67 (13.4%; 74.4%)</td>
</tr>
<tr>
<td>SIS-hand&lt;sup&gt;85&lt;/sup&gt;</td>
<td>9.4 (9.4%; NC)</td>
<td>21.9 (21.9%; NC)</td>
<td>25.9 (25.9%; NC)</td>
<td></td>
</tr>
<tr>
<td>SIS-hand-Brazilian Version</td>
<td>7.3 (7.4%; 29.2%)</td>
<td>17.1 (17.1%; 68.0%)</td>
<td>20.4 (20.4%; 81.1%)</td>
<td></td>
</tr>
<tr>
<td>TEMPA Time</td>
<td>38.4 (2.4%; 7.4%)</td>
<td>89.3 (5.5%; 17.2%)</td>
<td>106.4 (6.6%; 20.5%)</td>
<td></td>
</tr>
<tr>
<td>FAS</td>
<td>0.95 (3.5%; 12.3%)</td>
<td>2.2 (8.1%; 28.6%)</td>
<td>2.6 (10.0%; 33.8%)</td>
<td></td>
</tr>
<tr>
<td>Wolf Time</td>
<td>0.58-6.1 (0.5-5.1%; 16.0-26.5%)</td>
<td>1.3-14.2 (1.1-11.8%; 37.2-61.8%)</td>
<td>1.6-16.9 (1.3-14.1%; 44.2-73.4%)</td>
<td></td>
</tr>
<tr>
<td>Measure</td>
<td>SEM</td>
<td>MDC&lt;sub&gt;90&lt;/sub&gt;</td>
<td>MDC&lt;sub&gt;95&lt;/sub&gt;</td>
<td>LOA</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>--------</td>
</tr>
<tr>
<td>FAS</td>
<td>0.14-0.16</td>
<td>0.29-0.37</td>
<td>0.39-0.44</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.8-3.2%;</td>
<td>(5.8-7.4%;</td>
<td>(7.8-16%;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.0-5.6%)</td>
<td>11.5-11.6%)</td>
<td>13.7-15.6%)</td>
<td></td>
</tr>
</tbody>
</table>

* Only limits in the positive direction (from mean difference) were used to calculate % values; SEM: standard error of measurement; MDC: minimal detectable change; LOA: limits of agreement; %max: % of scale maximum score; %max: % of sample mean; NC: not calculated as mean not provided; AMAT: Arm Motor Ability Test; FAS: functional ability scale; QOM: quality of movement scale; ARAT: Action Research Arm Test; CAHAI: Chedoke Arm and Hand Activity Inventory; Duruoz: Duruoz Hand Index; MAL: Motor Activity Log; SIS: Stroke Impact Scale; TEMPA: Upper Extremity Performance Test for the Elderly; Wolf: Wolf Motor Function Test
Appendix E: Focus group schedules

Healthcare Providers

1. If you were to make a 2 level scale that consisted of high use and low use, how would you distinguish between the two levels?

Probe: What kinds of things would indicate someone is in the low use category or high use category? Any other ways to distinguish between low and high use?

2. If you were to make a 5 level scale where the highest level is “uses the arm the same as before the stroke” and the lowest level is unable to use their arm, what would the other three levels look like? Let us assume the dominant side is affected

Probes: Consider one level up from the lowest? Consider one level down from the highest? Should we consider quality of movement? Amount arm is used? Should we consider how much arms are used bilaterally? Unilaterally?

3. Think of your last 5 or so clients. Do these categories cover the range of use you have observed?

Probes: Do we need more categories? less categories? How many levels do you think would cover the full range of use?

4. So far we have assumed the dominant side was affected. Should the scale be the same if the non-dominant side was affected?

Client/Caregiver

1. Can you describe how you now use your stroke affected arm and hand in your daily life?

Probes: Can you describe situations when you use your stroke affected arm? Some activities people mentioned they have adapted by using only one hand, are there any tasks that people need to do with their affected hand no matter what? Can you describe situations when you need help with your daily tasks that require your hands? Can you describe tasks that you no longer perform due to the change in your arm and hand function? Can you describe how frequently use your stroke affected arm or hand? For caregiver: what kind of assistance do you provide for the arm and hand?
2. Let us now imagine there is a new treatment that could improve your stroke affected arm by a small amount. What would those small improvements that are important to you, look like?

Probes: What amount of improvement in your arm would you need to see before you start using your stroke affected hand a little more instead of only using your non-affected arm or hand? For caregivers: how would small improvements change the type of assistance you provide?
Appendix F: National Institute of Health Stroke Scale

Administrator stroke scale items in the order listed. Record performance in each category after each subscale exam. Do not go back and change scores. Follow directions provided for each exam technique. Scores should reflect what the patient does, not what the clinician thinks the patient can do. The clinician should record answers while administering the exam and work quickly. Except where indicated, the patient should not be coached (i.e. repeated requests to patient to make a special effort).

<table>
<thead>
<tr>
<th>Instructions</th>
<th>Scale Definition</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a. Level of Consciousness: The investigator must choose a response, even if a full evaluation is prevented by such obstacles as an endotracheal tube, language barrier, oro-tracheal trauma/bandages. A 3 is scored only if the patient makes no movement (other than reflexive posturing in response to noxious stimulation.</td>
<td>0 = Alert; keenly responsive. 1 = Not alert, but arousable by minor stimulation to obey, answer or respond. 2 = Not alert, requires repeated stimulation to attend, or is obtunded and requires strong or painful stimulation to make movements (not stereotyped). 3 = Responds only with reflex motor or autonomic effects or totally unresponsive, flaccid, areflexic.</td>
<td></td>
</tr>
<tr>
<td>1b. LOC Questions: The patient is asked the month and his/her age. The answer must be correct – there is no partial credit for being close. Aphasic and stuporous patients who do not comprehend the questions will score 2. Patients unable to speak because of endotracheal intubation, oro-tracheal trauma, severe dysarthria from any cause, language barrier, or any other problem not secondary to aphasia are given a 1. It is important that only the initial answer be graded and that the examiner not “help” the patient with verbal or nonverbal cues.</td>
<td>0 = Answers both questions correctly 1 = Answers one question correctly 2 = Answers neither question correctly</td>
<td></td>
</tr>
<tr>
<td>Instructions</td>
<td>Scale Definition</td>
<td>Score</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------</td>
<td>-------</td>
</tr>
<tr>
<td><strong>1c. LOC Commands:</strong> The patient is asked to open and close the eyes and then to grip and release the nonparetic hand. Substitute another one-step command if the hands cannot be used. Credit is given if an unequivocal attempt is made but not completed due to weakness. If the patient does not respond to command, the task should be demonstrated to them (pantomime) and score the result (i.e. follows none, one, or two commands). Patients with trauma, amputation, or other physical impediments should be given suitable one-step commands. Only the first attempt is scored.</td>
<td>0 = Performs both tasks correctly 1 = Performs one task correctly 2 = Performs neither task correctly</td>
<td></td>
</tr>
<tr>
<td><strong>2. Best Gaze:</strong> Only horizontal eye movements will be tested. Voluntary or reflexive (oculocephalic) eye movements will be scored but caloric testing is not done. If the patient has a conjugate deviation of the eyes that can be overcome by voluntary or reflexive activity, the score will be 1. If a patient has an isolated peripheral nerve paresis (CN III, IV, or V), score a 1. Gaze is testable in all aphasic patients. Patients with ocular trauma, bandages, preexisting blindness, or other disorder of visual acuity or fields should be tested with reflexive movements and a choice made by the investigator. Establishing eye contact and then moving about the patient from side to side will occasionally clarify the presence of a gaze palsy.</td>
<td>0 = Normal 1 = Partial gaze palsy. This score is given when gaze is abnormal in one or both eyes, but where forced deviation or total gaze paresis is not present. 2 = Forced deviation or total gaze paresis not overcome by the oculocephalic maneuver.</td>
<td></td>
</tr>
<tr>
<td><strong>3. Visual:</strong> Visual fields (upper and lower quadrants) are tested by confrontation, using finger counting or visual threat as appropriate. Patient must be encouraged, but if he or she looks at the side of the moving fingers appropriately, this can be scored as normal. If there is unilateral blindness or enucleation, visual fields in the remaining eye are scored. Score 1 only if a clear-cut asymmetry, including quadrantopia, is found. If patient is blind from any cause, score 3. Double stimulation is performed at this point. If there is extinction, patient receives a 1 and the results are used to answer question 11.</td>
<td>0 = No visual loss 1 = Partial hemianopia 2 = Complete hemianopia 3 = Bilateral hemianopia (blind including cortical blindness)</td>
<td></td>
</tr>
</tbody>
</table>
### Instructions

<table>
<thead>
<tr>
<th>Score</th>
<th>Scale Definition</th>
<th>Instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Normal symmetrical movement</td>
<td><strong>4. Facial Palsy:</strong> Ask or use pantomime to encourage the patient to show teeth or smile and close eyes. Score symmetry of grimace in response to noxious stimuli in the poorly responsive or noncomprehending patient. If facial trauma/bandages, orotracheal tube, tape, or other physical barrier obscures the face, these should be removed to the extent possible.</td>
</tr>
<tr>
<td>1</td>
<td>Minor paralysis (flattened nasolabial fold)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Partial paralysis (total or near total paralysis of lower face)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Complete paralysis (absence of facial movement in the upper and lower face)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>No movement</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Amputation, joint fusion; explain:</td>
<td><strong>5 and 6. Motor Arm and Leg:</strong> The limb is placed in the appropriate position; extend the arms $90^\circ$ (if sitting) or $45^\circ$ (if supine) and the leg $30^\circ$ (always tested supine). Drift is scored if the arm falls before 10 seconds or the leg before 5 seconds. The aphasic patient is encouraged using urgency in the voice and pantomime but not noxious stimulation. Each limb is tested in turn, beginning with the nonparetic arm. Only in the case of amputation or joint fusion at the shoulder or hip may the score be 9, and the examiner must clearly write the explanation for scoring as as a 9.</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5a</td>
<td>Left arm</td>
<td></td>
</tr>
<tr>
<td>5b</td>
<td>Right arm</td>
<td></td>
</tr>
</tbody>
</table>
### Instructions

<table>
<thead>
<tr>
<th>Score</th>
<th>Scale Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No drift, leg holds 30° position for full 5 seconds.</td>
</tr>
<tr>
<td>1</td>
<td>Drift; leg falls by the end of the 5-second period but does not hit bed</td>
</tr>
<tr>
<td>2</td>
<td>Some effort against gravity; leg falls to bed by 5 seconds but has some effort against gravity.</td>
</tr>
<tr>
<td>3</td>
<td>No effort against gravity; leg falls to bed immediately.</td>
</tr>
<tr>
<td>4</td>
<td>No movement</td>
</tr>
</tbody>
</table>
| 9     | Amputation, joint fusion; explain: _______

#### 6a

<table>
<thead>
<tr>
<th>Score</th>
<th>Scale Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left leg</td>
<td></td>
</tr>
</tbody>
</table>

#### 6b

<table>
<thead>
<tr>
<th>Score</th>
<th>Scale Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right leg</td>
<td></td>
</tr>
</tbody>
</table>

### 7. Limb Ataxia

This item is aimed at finding evidence of a unilateral cerebellar lesion. Test with eyes open. In case of visual defect, ensure testing is done in intact visual field. The finger-nose-finger and heel-shin tests are performed on both sides, and ataxia is scored only if present out of proportion to weakness. Ataxia is absent in the patient who cannot understand or is hemiplegic. Only in the case of amputation or joint fusion may the item be scored 9, and the examiner must clearly write the explanation for not scoring. In case of blindness, test by touching nose from extended arm position.

<table>
<thead>
<tr>
<th>Score</th>
<th>Scale Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Absent</td>
</tr>
<tr>
<td>1</td>
<td>Present in one limb</td>
</tr>
<tr>
<td>2</td>
<td>Present in both limbs</td>
</tr>
</tbody>
</table>

If present, is ataxia in

Right arm: 1 = Yes 2 = No
9 = Amputation or joint fusion; explain _______

Left arm: 1 = Yes 2 = No
9 = Amputation or joint fusion; explain _______

Right leg: 1 = Yes 2 = No
9 = Amputation or joint fusion; explain _______

Left leg: 1 = Yes 2 = No
9 = Amputation or joint fusion; explain _______
### Instructions

**8. Sensory:** Sensation or grimace to pin prick when tested or withdrawal from noxious stimulus in the obtunded or aphasic patient. Only sensory loss attributed to stroke is scored as abnormal, and the examiner should test as many body areas (arms (not hands), legs, trunk, face) as needed to accurately check for hemisensory loss. A score of 2, “severe or total,” should only be given when a severe or total loss of sensation can be clearly demonstrated. Stuporous and aphasic patients will therefore probably score 1 or 0. The patient with brain stem stroke who has bilateral loss of sensation is scored 2. If the patient does not respond and is quadriplegic, score 2. Patients in coma (question 1a=3) are arbitrarily given a 2 on this item.

<table>
<thead>
<tr>
<th>Scale Definition</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = Normal; no sensory loss</td>
<td></td>
</tr>
<tr>
<td>1 = Mild to moderate sensory loss; patient feels pin prick is less sharp or is dull on the affected side, or there is a loss of superficial pain with pin prick but patient is aware he or she is being touched.</td>
<td></td>
</tr>
<tr>
<td>2 = Severe to total sensory loss; patient is not aware of being touched.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Instructions</th>
<th>Scale Definition</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 = No aphasia, normal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 = Mild to moderate aphasia; some obvious loss of fluency or facility of comprehension, without significant limitation on ideas expressed or form of expression. Reduction of speech and/or comprehension, however, makes conversation about provided material difficult or impossible. For example, in conversation about provided materials examiner can identify picture or naming card from patient’s response.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 = Severe aphasia; all communication is through fragmentary expression; great need for inference, questioning, and guessing by the listener. Range of information that can be exchanged is limited; listener carries burden of communication. Examiner cannot identify materials provided from patient’s response.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 = Mute, global aphasia, no usable speech or auditory comprehension.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Best Language: A great deal of information about comprehension will be obtained during the preceding sections of the examination. The patient is asked to describe what is happening in the attached picture, to name the items on the attached naming sheet, and to read from the attached list of sentences. Comprehension is judged from responses here as well as to all of the commands in the preceding general neurological exam. If visual loss interferes with the tests, ask the patient to identify objects placed in the hand, repeat, and produce speech. The intubated patient should be asked to write sentence. The patient in coma (question 1a = 3) will arbitrarily score 3 on this item. The examiner must choose a score in the patient with stupor or limited cooperation, but a score of 3 should be used only if the patient is mute and follows no one-step commands.
<table>
<thead>
<tr>
<th>Instructions</th>
<th>Scale Definition</th>
<th>Score</th>
</tr>
</thead>
</table>
| **10. Dysarthria:** If a patient is thought to be normal an adequate sample of speech must be obtained by asking patient to read or repeat words from the attached list. If the patient has severe aphasia, the clarity of articulation of spontaneous speech can be rated. Only if the patient is intubated or has other physical barrier to producing speech, may be item be scored “9” and the examiner must clearly write an explanation for not scoring. Do not tell the patient why he/she is being tested. | 0 = Normal  
1 = Mild to moderate; patient slurs at least some words and, at worst, can be understood with some difficulty.  
2 = Severe; patient’s speech is so slurred as to be unintelligible in the absence of or out of proportion to any dysphasia, or is mute/anarthric.  
9 = Intubated or other physical barrier, explain | |
| **11. Extinction and Inattentioin (formerly Neglect):** Sufficient information to identify neglect may be obtained during the prior testing. If the patient has a severe visual loss preventing visual double simultaneous stimulation, and the cutaneous stimuli are normal, the score is normal. If the patient has aphasia but does not appear to attend to both sides, the score is normal. The presence of visual spatial neglect or anosognosia may also be taken as evidence of abnormality. Since the abnormality is scored only if present, the item is never untestable. | 0 = No abnormality  
1 = Visual, tactile, auditory, spatial, or personal inattention or extinction to bilateral simultaneous stimulation in one of the sensory modalities.  
2 = Profound hemi-inattention or hemi-inattention to more than one modality. Does not recognize own hand or orients to only one side of space. | |
Appendix G: Montreal Cognitive Assessment

![Montreal Cognitive Assessment Diagram](image)

**Visual-Spatial / Executive**
- Copy cube
- Draw clock (10 past eleven), score 3 points

**Naming**
- Rhinoceros
- Elephant
- Camel

**Memory**
- Read list of words, subject must repeat them. Do 2 trials. Do a recall after 5 minutes.
- 1st trial:
  - Face
  - Velvet
  - Church
  - Daisy
  - Red
- 2nd trial:
  - Face
  - Velvet
  - Church
  - Daisy
  - Red

**Attention**
- Read list of digits (1 digit/sec). Subject has to repeat them in the forward order and backward order.
- Forward order: 2 1 8 5 4
- Backward order: 7 4 2

**Language**
- Repeat: I only know that John is the one to help today.
- The cat always hid under the couch when dogs were in the room.

**Abstraction**
- Similarity between e.g., banana - orange = fruit • train - bicycle = vehicle
- Watch - rule

**Delayed Recall**
- Has to recall words with no cue:
  - Face
  - Velvet
  - Church
  - Daisy
  - Red
  - Points for uncued recall only

**Orientation**
- Optional:
  - Category cue
  - Multiple choice cue

**Total**
- Date
- Month
- Year
- Day
- Place
- City
- Normal ≥ 26 / 30

Add 1 point if ≤ 12yr olds
Appendix H: Motor Activity Log 14

<table>
<thead>
<tr>
<th>Items</th>
<th>During the last week, did you use your weaker arm to [state activity]?</th>
<th>Can you tell me why it wasn’t used, using this list of possible reasons?</th>
<th>Can you tell me how much you used your weaker arm to [state activity]? So you believe you used your arm [restate use description]. Is that correct?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hold a book/magazine/newspaper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Use a towel to dry self</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Pick up a glass (doesn’t include drinking)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Brush teeth (doesn’t include preparing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Shave/Put on Make-up (or lotion)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Open door with a key</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Write/type (if non-dominant hand was affected ask about typing)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Steady self</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Put arm through clothing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Carry object</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Grasp fork/spoon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Comb hair</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Pick up a cup by a handle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Button clothes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix I: Action Research Arm Test

The ARAT is a measure to assess specific changes in upper limb function among people who have sustained brain injuries resulting in weakness on one side. It assesses a client's ability to handle objects differing in size, weight and shape.

That ARAT consists of 19 items that are grouped into four subtests:

1. Grasp
2. Grip
3. Pinch
4. Gross movement

All items are rated on a 4-point ordinal scale ranging from 0 (no movement possible) to 3 (normal performance of the task). The subtest scores vary according to the number of items performed in each subtest.

The total score on the ARAT ranges from 0-57, with a higher score indicating better performance.

Instructions:

Test on the stroke affected arm only. Subtests are ordered in such a way that if the person scores 3 on item one (the most difficult) the person is credited with having scored 3 on all items of the subtest. You don’t have to test the remaining subtest items. Each task should be demonstrated. You can ask the client to perform the task again if you thought they didn’t understand the instructions.

If the person scores less than 3 on item one, then item two is administered. Item two is the easiest item in each of the subtests and if the person scores 0, then he/she is given a 0 for the remaining subtests. Move to the next subtest.

If the person scores less than 3 on item one and more than 0 on item two, all items in the subtest must be administered.

Scoring Scale:

3: Performs test normally
2: Completes test, but takes abnormally long time or uses compensations
1: Performs test partially (e.g. grasps and releases object but cannot reach the shelf)
0: Can perform no part of test (e.g. cannot grasp the object)
<table>
<thead>
<tr>
<th><strong>GRASP</strong></th>
<th><strong>Evaluation</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left</strong></td>
<td><strong>Right</strong></td>
<td><strong>/18</strong></td>
</tr>
<tr>
<td>1. Block 10cm (if score = 3, total = 18 go to Grip test)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Block 2.5cm (if score = 0, rest = 0 go to Grip test)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Block 5cm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Block 7.5cm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Ball 7.5cm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Stone 10 x25 x 1 cm</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subtotal GRASP</strong></td>
<td><strong>/18</strong></td>
<td><strong>/18</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>GRIP</strong></th>
<th><strong>Evaluation</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left</strong></td>
<td><strong>Right</strong></td>
<td><strong>/12</strong></td>
</tr>
<tr>
<td>1. Pour water glass to glass (if score = 3, total = 12 ; go to Pinch test)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Move the 2.25cm tube a distance of 30 cm (if score = 0, total is 0, go to Pinch test)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Move the 1cm tube as above</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Put washer over a bolt</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subtotal Grip</strong></td>
<td><strong>/12</strong></td>
<td><strong>/12</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>PINCH</strong></th>
<th><strong>Evaluation</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left</strong></td>
<td><strong>Right</strong></td>
<td><strong>/18</strong></td>
</tr>
<tr>
<td>1. Ball bearing of 6mm picked up between thumb and ring finger (if score = 3, total = 18 and go to Gross movement test)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Marble picked up between thumb and index finger (if score = 0, total = 0, and go to Gross movement test)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Ball bearing 6mm picked up between thumb and middle finger</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Ball bearing of 6mm picked up between thumb and index finger</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Marble picked up between thumb and ring finger</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Marble picked up between thumb and middle finger</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SUBTOTAL Pinch</strong></td>
<td><strong>/18</strong></td>
<td><strong>/18</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Gross Movement</strong></th>
<th><strong>Evaluation</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Left</strong></td>
<td><strong>Right</strong></td>
<td><strong>/9</strong></td>
</tr>
<tr>
<td>1. Place hand behind head (if score = 3, total = 9; if score = 0, total = 0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Place hand on top of head</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Touch mouth with hand</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SUBTOTAL Gross Movement</strong></td>
<td><strong>/9</strong></td>
<td><strong>/9</strong></td>
</tr>
</tbody>
</table>
Appendix J: Stroke Impact Scale-hand

The purpose of this questionnaire is to assess participants’ perceptions of their hand difficulties.

**Instructions for participants:** The following questions are about your ability to use your hand that was MOST AFFECTED by your stroke.

*In the past 2 weeks, how difficult was it to use your hand that was most affected by your stroke to……*

<table>
<thead>
<tr>
<th></th>
<th>Not difficult at all</th>
<th>A little difficult</th>
<th>Somewhat difficult</th>
<th>Very difficult</th>
<th>Could not do at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. carry heavy objects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e.g. bag of groceries)?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. turn a doorknob?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. open a can or jar?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. tie a shoe lace?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. pick up a dime?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix K: Chedoke-McMaster Stroke Assessment (arm, hand and shoulder scales)

**ARM and HAND:** Start at stage 3. Starting positions: sitting with forearm in lap in a neutral position, wrist at 0° and fingers slightly fixed. Changes from this position are indicated by underlining. Place an X in the box of each task accomplished. Score the highest stage in which the client achieves at least two Xs.

**ARM**

1  □ not yet stage 2

2  □ resistance to passive shoulder abduction or elbow extension
   □ facilitated elbow extension
   □ facilitated elbow flexion

3  □ touch opposite knee
   □ touch chin
   □ shoulder shrugging >1/2 range

4  □ extension synergy, then flexion synergy
   □ shoulder flexion to 90°
   □ elbow at side, 90° flexion: supination, then pronation

5  □ flexion synergy, then extension synergy
   □ shoulder abduction to 90° with pronation
   □ shoulder flexion to 90°: pronation then supination

6  □ hand from knee to forehead 5x in 5sec
   □ shoulder flexion to 90°:trace a figure 8
   □ arm resting at side of body: raise arm overhead with full supination

7  □ clap hand overhead, then behind back 3x in 5 sec
   □ shoulder flexion to 90°: scissor in front 3x in 5 sec
   □ elbow at side, 90° flexion*: resisted shoulder external rotation

**STAGE OF ARM _____________**
HAND

1  □ not yet stage 2

2  □ positive Hoffman
   □ resistance to passive wrist or finger extension facilitated finger flexion
   □ facilitated finger flexion

3  □ wrist extension > ½ range
   □ finger/wrist flexion > ½ range
   □ supination, thumb in extension: thumb to index finger

4  □ finger extension, then flexion
   □ thumb extension > ½ range, then lateral prehension
   □ finger extension with lateral prehension

5  □ finger flexion, then extension
   □ pronation: finger abduction
   □ hand unsupported: opposition of thumb to little finger

6  □ pronation: tap index finger 10x in 5 sec
   □ pistol grip: pull trigger, then return
   □ pronation: wrist and finger extension with finger abduction

7  □ thumb to finger tips, then reverse 3x in 12 sec
   □ bounce a ball 4 times in succession, then catch
   □ pour 250 ml. from 1 litre pitcher, then reverse

STAGE OF HAND _______________
SHOULDER PAIN

☐ 1. Constant, severe arm and shoulder pain with pain pathology in more than just the shoulder

☐ 2. Intermittent, severe arm and shoulder pain with pain pathology in more than just the shoulder

☐ 3. Constant shoulder pain with pain pathology in just the shoulder

☐ 4. Intermittent shoulder pain with pain pathology in just the shoulder

☐ 5. Shoulder pain is noted during testing, but the functional activities that the client normally performs are not affected by the pain.

☐ 6. No shoulder pain, but at least one prognostic indicator is present
   - Arm stage 1 or 2
   - Scapula malaligned
   - Loss of range of shoulder movement: flexion/abduction <90° or external rotation <60°

☐ 7. Shoulder pain and prognostic indicators are absent

STAGE OF SHOULDER PAIN ________________