

Health Related Quality of Life over One Year Post Stroke:

Identifying Response Shift Susceptible Constructs

by

Ruth Barclay-Goddard

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Department of Community Health Sciences

University of Manitoba

Winnipeg

Copyright © 2008 Ruth Barclay-Goddard

Abstract

Problem: Many individuals with chronic illnesses such as stroke and ongoing activity limitations report self-perceived health related quality of life (HRQL) that is similar to that of healthy individuals. This phenomenon is termed response shift (RS). RS describes how people change: internal standards in assessing HRQL (recalibration), values (reprioritization), or how they define HRQL (reconceptualization), after an event such as stroke. Changes in HRQL post stroke may be inaccurate if RS is not taken into account. Increased knowledge of RS may affect the way in which HRQL measures are used, both clinically and in research. The overall objective was to assess RS in construct specific HRQL models post stroke: physical function, mental health, and participation.

Methods: Data were analysed from the longitudinal study “Understanding Quality of Life Post-Stroke: A Study of Individuals and their Caregivers”. Six-hundred and seventy-eight persons with stroke at 1, 3, 6, and 12 months post stroke participated. Generic and stroke specific HRQL measures were collected. Descriptive analysis was completed with SAS, and identification of RS utilized structural equation modeling with LISREL.

Results: Mean age of participants was 67 years (SD 14.8), and 45% were female. RS was identified in mental health using a framework which was developed for identifying RS statistically with multiple time points. RS was also identified in physical function where it had not been expected, possibly due to the self perceived nature of the response options. The effect size of change in physical function was affected by the presence of RS. The timing of RS in mental health and physical function was primarily around the 12

month time period, and predominantly recalibration RS. RS was also identified in participation.

Conclusions: The framework that was developed was useful in identifying RS and incorporated important issues such as multiple testing and validation of the model. The presence of RS affects measurement of HRQL constructs post stroke; recalibration RS can be measured clinically with specific methods to account for RS. RS should also be measured in research studies to ensure accurate measurement of change. Future research should evaluate additional models in stroke and other populations.

Acknowledgments

I would like to especially thank my family: Perry, Bethany, and Hannah as well as my mother who have believed in me throughout the entire PhD experience – I could not have done it without them.

I would also like to thank my PhD Committee. Dr. Jure Manfreda was my initial advisor prior to retirement. Thanks to Dr. Nancy Mayo, my advisor, for believing in me and taking me on with an impossibly busy schedule, taking time for me during visits to McGill, and for providing expertise and positive, constructive feedback. Dr. Lisa Lix provided expertise in longitudinal data analysis and structural equation modeling, and support and encouragement as a ‘local advisor in Winnipeg’. Dr. Bob Tate supplied valuable statistical expertise and insightful advice, while Dr. Leah Weinberg provided helpful feedback on clarity and relevance to rehabilitation. Dr. Sharon Wood-Dauphinee was the principal investigator of the study which was analysed, giving permission to use the data. I am also grateful that both Dr. Wood Dauphinee and Dr. Mayo made funding available for me to travel occasionally to McGill for invaluable face to face meetings.

Thanks also to Susan Scott at McGill for clarifying many data questions for me, and Phil Gerson in Psychology at University of Manitoba for lending me a much needed LISREL manual at a crucial time.

Thanks to colleagues for support and to Jenneth Swinamer for flexibility of work hours.

I received a University of Manitoba Graduate Fellowship for 2002-3, and 2003-4, and UMSU scholarship 2002-3, both of which were greatly appreciated.

Dedication

I would like to dedicate this thesis to my father, David Barclay, who saw me embark on the PhD process, but was unable to see me finish. I would also like to dedicate this to Perry, my husband and Bethany and Hannah, my daughters, as well as my mother who have supported and believed in me through everything.

Table of Contents

List of Tables	x
List of Figures	xi
Preface	xii
Chapter 1: Introduction	1
Quality of Life and Health Related Quality of Life	1
Types of change and definition of response shift:	4
Stroke as a model of response shift	6
Table 1-1 The International Classification of Functioning (ICF) ³⁷	8
Figure 1-1 Association between the ICF domains and the constructs of HRQL	9
HRQL measurement post stroke	10
Response shift post stroke	12
SEM and Measurement Invariance	14
Implications of Response Shift Measurement - Why is it important to study Response Shift?	17
Reference List	19
Chapter 2: Much research needs to be done on Response Shift in order to accurately evaluate change in patient reported outcomes (PROs)	29
Abstract	30
Introduction	33
What is Response Shift?	34
Theoretical Models of Response Shift	36
Study Design to Address Response Shift	38
Retrospective Ratings	38
Individualized Methods	39
Direct Questioning	40
Vignettes Ratings	41
Questionnaire	41
Statistical Methods to Address Response Shift	42
Growth curve analysis	44
Residual Analysis	44
Multivariate multi-level models	46
Rasch analysis	46
Convergent validity	47
Clinical importance	48
Recommendations for Future Research Priorities	49
Further evaluation of Current Methods and Investigation of Newer Methods	50
Clinical importance	52
Reporting guidelines	53
Conclusion	53
Acknowledgements	54
Reference List	55
Figure 2-1 Then test	64
Figure 2-2 The Patient Generated Index (PGI)	65
Figure 2-3 Factor Analysis: A Hypothetical Measurement Model of HRQL	66

Table 2- 1 Summary of Recommendations of Future Research Priorities	67
Chapter 3: Study Methodology and Descriptive Data Analysis	68
Summary of chapter 3	68
Rationale for the thesis.....	68
Global Aim.....	70
Overall Objective	70
Hypotheses	71
Source of data	71
Table 3-1 Data Source Variables	72
Figure 3-1 Association of the ICF , HRQL constructs and HRQL indices.....	73
Table 3-2 Items and subscales from HRQL measures and related constructs.....	74
Analytical Approach	74
Table 3-3 The Analytical Approach	75
Analyses for missing data	75
Characterizing variables (non-normality).....	76
Latent curve models.....	77
Building Models of the HRQL constructs	77
Determining response shift	78
Results.....	79
Characteristics of persons	79
Pattern of missing data.....	79
Table 3-4 Factors Associated with Missingness.....	80
Characteristics of the variables	80
Latent curve models.....	81
Table 3-5 Latent Curve Models Suggesting Linear Change.....	82
Supplemental Tables.....	83
Supplemental Table 3-1 HRQL Measures.....	83
Supplemental Table 3-2 Description of the Cohort.....	87
Supplemental Table 3-3 Outcomes at all Times	88
Supplemental Table 3-4 Loss to Follow-up.....	91
Supplemental Table 3-5 Testing for Normality in HRQL Indices	92
Following Chapters.....	93
Reference List.....	94
Chapter 4: Linking Literature and Methods to a Statistical Framework for Examining Response Shift	101
Chapter 5: Response Shift was Identified in Health Related Quality of Life Using a Framework Developed for Multiple Time Points: A Demonstration with Individuals Post Stroke	103
Introduction.....	107
Structural Equation Modeling and Response Shift.....	108
Development of a Framework to Test Response Shift	109
Application of Framework.....	112
Description of Example Dataset	112
Preparatory work to develop the model.....	113
Step 1 – Identifying the Presence of Response Shift	115
Step 2 - Validation of the model.....	115

Step 3 Identify Localized Sources(s) of Response Shift.....	115
Step 4 - Exploratory Approach – Test Sources that are “Theory Supported”	115
Step 4 – Theory-driven Approach – ‘a priori’	116
Discussion.....	116
Conclusions.....	118
Figure 5-1 Identifying Response Shift with Multiple Time Points.....	120
Figure 5-2 Mental Health Model – 1 month.....	121
Table 5-1 HRQL Indices and Variables Utilized in the Mental Health Model ..	122
Table 5-2 Description of Study Participants.....	123
Table 5-3 Summary Statistics for Mental Health Variables at Four Measurement Occasions Post Stroke.....	124
Table 5-4 Localization of Response Shift: Changes in Non-Standardized Parameter Estimates.....	125
Appendix 5-1 Mental Health Model Validity - Sensitivity Analysis	126
Chapter 7: Can Response Shift Occur in Self-Perceived Physical Function? Using Stroke as a Model.....	135
Introduction.....	138
Methods and Data analysis	141
Results.....	144
Discussion.....	147
Conclusion	149
Figure 7-1 Physical Function Model – 1 month	151
Table 7-1 Items and Subscales from HRQL Indices and Associated Latent Variables	152
Table 7-2 Demographics of Study Participants	153
Table 7-3 Outcomes of Measured Variables over One Year Post Stroke.....	154
Table 7-4 Localization of Response Shift - Changes in Non-Standardized Parameter Estimates.....	155
Table 7-5 Effect Sizes of Observed Change, Response Shift, and True Change	156
Appendix 7-1 Physical Function Model Validity (sensitivity analysis) for Longitudinal Model	157
References.....	158
Chapter 8: The Participation Model.....	165
Link to Chapter 8	165
Participation	166
Methods.....	167
Table 8-1 Items and subscales from HRQL indices and related constructs	168
Results.....	168
Table 8-2 Outcomes of HRQL Indices in Participation Models at all Times.....	169
The measurement model.....	170
Table 8-3 Cross-sectional models – Participation	171
Figure 8-1 Participation Model – 1 month.....	171
Figure 8-2 Participation Model – 12 months	172
Conclusion	175
Reference List.....	176
Chapter 9: Conclusion.....	179

Summary of Hypotheses 179

 Hypothesis 1: 179

 Hypothesis 2: 180

 Hypothesis 3: 181

 Hypothesis 4: 181

Overall Summary 182

The benefits of identifying the timing, type, and measured variables associated with
response shift 183

 Timing 183

 Table 9-1 Response Shift Localization 184

 Type of response shift 185

 Measured Variable 186

 Table 9-2 Future research based on thesis work 188

Overall limitations 189

Implications 190

Reference List 193

List of Tables

Table 1-1 The International Classification of Functioning (ICF) ³⁷	8
Table 2-1 Summary of Recommendations of Future Research Priorities	67
Table 3-1 Data Source Variables	72
Table 3-2 Items and subscales from HRQL measures and related constructs	74
Table 3-3 The Analytical Approach	75
Table 3-4 Factors Associated with Missingness	80
Table 3-5 Latent Curve Models Suggesting Linear Change	82
Supplemental Table 3-1 HRQL Measures	83
Supplemental Table 3-2 Description of the Cohort	87
Supplemental Table 3-3 Outcomes at all Times	88
Supplemental Table 3-4 Loss to Follow-up	91
Supplemental Table 3-5 Testing for Normality in HRQL Indices	92
Table 5-1 HRQL Indices and Variables Utilized in the Mental Health Model ..	122
Table 5-2 Description of Study Participants	123
Table 5-3 Summary Statistics for Mental Health Variables at Four Measurement Occasions Post Stroke	124
Table 5-4 Localization of Response Shift: Changes in Non-Standardized Parameter Estimates	125
Appendix 5-1 Mental Health Model Validity - Sensitivity Analysis	126
Table 7-1 Items and Subscales from HRQL Indices and Associated Latent Variables	152
Table 7-2 Demographics of Study Participants	153
Table 7-3 Outcomes of Measured Variables over One Year Post Stroke	154
Table 7-4 Localization of Response Shift - Changes in Non-Standardized Parameter Estimates	155
Table 7-5 Effect Sizes of Observed Change, Response Shift, and True Change	156
Appendix 7-1 Physical Function Model Validity (sensitivity analysis) for Longitudinal Model	157
Table 8-1 Items and subscales from HRQL indices and related constructs	168
Table 8-2 Outcomes of HRQL Indices in Participation Models at all Times	169
Table 8-3 Cross-sectional models – Participation	171
Table 9-1 Response Shift Localization	184
Table 9-2 Future research based on thesis work	188

List of Figures

Figure 1-1 Association between the ICF domains and the constructs of HRQL....	9
Figure 2-1 Then test.....	64
Figure 2-2 The Patient Generated Index (PGI).....	65
Figure 2-3 Factor Analysis: A Hypothetical Measurement Model of HRQL	66
Figure 3-1	73
Figure 5-1 Identifying Response Shift with Multiple Time Points.....	120
Figure 5-2 Mental Health Model – 1 month	121
Figure 7-1 Physical Function Model – 1 month	151
Figure 8-1 Participation Model – 1 month.....	171
Figure 8-2 Participation Model – 12 months	172

Preface

As a physiotherapist with a clinical background and interest in stroke rehabilitation, I became interested in outcome measurement post stroke; this became a focus of my Masters work at McMaster University in 1993-4. I became interested in response shift after an enlightening discussion with Dr. Nancy Mayo in 2004 about measuring health related quality of life (HRQL) post stroke and the phenomenon of response shift. At that point, I was interested in doing PhD thesis work focusing on HRQL post stroke, but was unfamiliar with response shift. Dr. Mayo's PhD student at the time, Sara Ahmed, had just completed thesis work on response shift post stroke; it was apparent that further research was required in this area. I knew nothing of structural equation modeling, a method used to identify response shift, but was intrigued by the chance to learn more. Hence, my initiation to the concept of response shift with SEM led to this thesis, Health Related Quality of Life over One Year Post Stroke: Identifying Response Shift Susceptible Constructs.

Organization of the Thesis

This thesis incorporates three manuscripts which will be or have been peer reviewed. The thesis was successfully defended on August 15, 2008. The composition of this thesis is based on the status of the manuscripts as of August 20, 2008. Two have been submitted: one is under review and comments for revisions have been received for the other. A third paper will be submitted in September 2008. The published manuscripts will be in a different format than in this thesis, after peer review. Readers are encouraged to refer to the final published manuscripts.

Chapter 1 provides an overview of quality of life (QOL) and HRQL, and the sequelae of stroke. In describing the impact of stroke we use the World Health Organization International Classification of Functioning and Disability (ICF) model. To understand response shift in HRQL requires understanding the ICF based components. This thesis therefore addresses components of HRQL that are susceptible to response shift. *Chapter 1* also provides an introduction to response shift and leads to a more detailed view of response shift presented in the manuscript in *Chapter 2*.

Manuscript 1 is presented in *Chapter 2*, titled “Much research needs to be done on Response Shift in order to accurately evaluate change in patient reported outcomes PROs”. The manuscript outlines a research agenda for response shift. This paper arose out of the first International Society of Quality of Life Research (ISOQOL) response shift special interest group (SIG) meeting in San Francisco in 2005 at which I along with Joshua Epstein were “commissioned” to set the stage for future research in response shift. I took the lead with input from Joshua and from my supervisor, Nancy Mayo. Contributions for the content were also provided by the leaders in the field of response shift.

Chapter 3 provides objectives and hypotheses for the entire thesis, description of the data source, analysis of data, and the characteristics of study participants and variables.

Chapter 4 provides a link between the introductory Chapters and the following manuscript.

Manuscript 2 is presented in *Chapter 5*, “Response Shift was Identified in Health Related Quality of Life Using a Framework Developed for Multiple Time Points: A

Demonstration with Individuals Post Stroke”. In this manuscript, a framework is described and demonstrated that can assist researchers in identifying response shift over time post stroke using a previously recognized method of SEM to determine presence of response shift. The mental health construct of HRQL was used to demonstrate the framework.

A description of the link between the two previous manuscripts and the following manuscript is presented in *Chapter 6*. A further demonstration of the identification of response shift susceptible constructs of HRQL is presented in *Manuscript 3* in *Chapter 7*, “Can Response Shift Occur in Self-Perceived Physical Function? Using Stroke as a Model”. Finding that response shift was present in physical function post stroke leads to some questions about measurement of physical function post stroke.

Chapter 8 presents a novel model of participation over one year post stroke. An initial determination of response shift over time is also presented in this chapter. The conclusions presented in *Chapter 9* puts together the results of all the studies and provides suggestions for future response shift studies.

Chapter 1: Introduction

Self report measures of health related quality of life (HRQL) are recognized as important components in the assessment of the impact of interventions, whether in a clinical or research setting. Self report measures are also referred to as patient reported outcomes (PROs). PROs are increasingly used to assess the impact of interventions including drugs, as evidenced by the United States Food and Drug Administration (FDA) guidance and European guidance for these measures ^{1,2}. As these measures are used more and more, it has been recognized that a phenomenon known as response shift will need to be considered when interpreting PRO results. Response shift describes the often paradoxical findings that occur when individuals with chronic disability or life threatening disease report that their HRQL is stable or similar to those who are healthy ³⁻⁸. This chapter will first review the concepts of QOL and HRQL, in the context of stroke, provide a description of response shift in HRQL measurement, and discuss how stroke can be used as a model for response shift. There will be a focus on the use of structural equation modeling to identify response shift along with a discussion of measurement invariance. To conclude, implications of response shift measurement will be described.

Quality of Life and Health Related Quality of Life

The terms quality of life (QOL) and health related quality of life (HRQL) are often vague and there has been little consensus on definition; various domains of HRQL have been described by different investigators ⁹⁻¹³. To make matters more confusing, the terms QOL and HRQL are often used interchangeably ¹⁴. Some authors state that they are estimating QOL but describe domains that are consistent with HRQL.

QOL is a broad concept which is self perceived, multidimensional, and reflected in cultural, social, and environmental contexts¹⁴. The concept of QOL incorporates all aspects of life and has been used in a variety of disciplines such as: geography, philosophy, medical sciences, social sciences, health promotion, and advertising¹². In the social sciences, QOL definitions often refer to “the adequacy of people’s material circumstances and to their feelings about these circumstances”¹³. The components of quality of life have been comprehensively identified by Flanagan as: i) health, personal safety, material well being; ii) having children and relationships with others (spouse, close friends, and relatives); iii) participation in public affairs, helping and encouraging others; iv) intellectual development, self-understanding, occupational role, and creative expression; v) leisure activities and socializing^{15, 16}. Other components have also been described: housing, perceptions of the environment, level of freedom, opportunity, cultural, and spiritual aspects^{12, 13}.

The World Health Organization (WHO) defines QOL as “An individual’s perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns. It is a broad ranging concept affected in a complex way by the person’s physical health, psychological state, personal beliefs, social relationships and their relationship to salient features of their environment”¹⁷. WHO therefore describes the following six domains of QOL: physical health, psychological state, level of independence, social relationships, environmental features, and spiritual concerns and personal beliefs¹⁸.

HRQL is “limited to the aspects of life that are important to the evaluator in the context of health and illness”¹⁴. Six HRQL domain frameworks have been summarized; all six

frameworks included physical, interpersonal (i.e. social aspects), and psychological / spiritual domains ¹⁴. Two included economic / environmental factors as part of HRQL, and a different two included a subjective general health perception component. The three domains commonly used to estimate HRQL include physical, psychological, and social functioning ¹⁹. There is not full agreement, however; Torrance has stated that HRQL consists of physical function and emotional function only and considered social functioning to be “beyond the skin”, not a part of HRQL ²⁰. There is a consensus that the impression of overall health / general well being should also be included in HRQL assessment ¹⁹. The benefits of measuring HRQL are that the scope of outcome measurement in health is broadened and that the client’s judgment may be able to influence treatment choices ¹³.

HRQL is self perceived (‘subjective’), multi-dimensional, can be affected by disease and by various forms of treatment, and it can change over time ^{11, 19, 21}. For the purpose of this thesis, the definition of HRQL utilized is “the value assigned to duration of life as modified by the impairments, functional states, perceptions and social opportunities influenced by disease, injury, treatment or policy” ²².

There are two views of HRQL measurement: positivist views and constructivist views. In the positivist view, individuals use some external reference to evaluate HRQL; a particular score is assumed to mean the same for all people with that score ²³.

Constructivist views assume that the judgment of the individual in evaluating HRQL is dependent on memory and how an individual reconstructs the event. Self assessed health is valid, as it represents the perceived state and is a representation of reality for the

individual. Differences in scores between and within individuals are real, reflecting both known and unknown factors (such as response shift) ²³.

A conceptual framework for clinical outcomes relating to HRQL has been described ²⁴. The Wilson and Cleary model includes relationships between physiologic variables, symptoms, functioning, and general health perception. It also includes characteristics of the individual and the environment and the relationship to quality of life. Over time, two or more aspects may vary together, at different rates, or in opposite directions ²⁴.

Types of change and definition of response shift:

To measure change in HRQL, the pre test score is typically subtracted from the post test score. This assumes that the underlying conceptualization and understanding of HRQL has not changed over time for the individual. As alluded to in the first paragraph, paradoxical findings may occur, where individuals with a chronic or life threatening condition may describe HRQL as similar to healthy individuals ³⁻⁸. When this occurs, the individual's internal standards, values, or conceptualizations of HRQL may have changed over time ²⁵. In this case, change scores may be incorrectly interpreted if response shift is not taken into account ^{4, 26, 27}.

Response shift was originally based in educational training, organizational change, and management science and more recently the concept has been utilized in HRQL research ^{3, 25, 28, 29}. An earlier description of changes in the conceptualization of a concept such as HRQL was described by Golembiewski et al in 1976. Three types of observed change in self report ratings were discussed: alpha, beta, and gamma change. "Alpha change involves a variation in the level of some existential state, given a constantly calibrated

measuring instrument related to a constant conceptual domain”³⁰. Alpha change is considered true behavioural change, which is typically measured with pretest – posttest study designs^{30, 31}.

“Beta change involves a variation in the level of some existential state, complicated by the fact that some intervals of the measurement continuum associated with a constant conceptual domain have been recalibrated”³⁰. With scale recalibration, the individual changes his / her internal measurement standards; it may be compared to a ruler that shrinks or stretches³¹.

“Gamma change involves a redefinition or reconceptualization of some domain, a major change in the perspective or frame of reference within which phenomena are perceived and classified, in what is taken to be relevant in some slice of reality”³⁰. With concept redefinition / reconceptualization, a specific construct (such as HRQL) is reconceptualized. On subsequent testing, the understanding of the construct being tested changes for the individual; the measures used at the pretest may no longer be appropriate at the posttest^{30, 31}. It has been suggested that if change is not reconceptualization or recalibration, then it must be true behavioural change³¹.

The various types of change, as described above, are also referred to as response shift. Response shift is defined as “a change in the meaning of one’s self-evaluation of a target construct as a result of: (a) a change in the respondent’s internal standards of measurement (i.e. scale recalibration); (b) a change in the respondent’s values (i.e. the importance of component domains constituting the target construct) or (c) a redefinition of the target construct (i.e. reconceptualization)”²⁵. A change in values (also called

reprioritization) was not described in the alpha, beta, gamma framework, but added by Schwartz and Sprangers in 1999^{3, 25}.

The various types of response shift are likely interdependent; they may occur together, in parallel, or at the same rate³². Recalibration / beta change may occur along with reconceptualization / gamma change or reprioritization. If the anchors for measurement change with a recalibration, perhaps a reconceptualization also occurs; if the response scale increases or decreases, then the meanings of the anchors are also likely changed^{25, 32}. It has been suggested that due to the necessity for more research in response shift, “it may be reasonable in some instances to limit one’s investigation to simply detecting response shift”³³. This will be discussed further in Chapter 9 (conclusion).

The need for further research into response shift was echoed by the response shift special interest group which led to the manuscript that makes up Chapter 2, “Much research needs to be done on Response Shift in order to accurately evaluate change in patient reported outcomes (PROs)”. A more detailed description and review of response shift theory, methodology, and research priorities are described in the manuscript.

Stroke as a model of response shift

Response shift is likely to occur if change in health is recent, intense and all-encompassing³. It has been suggested that recalibration is more likely in the first few months after a threatening event and that clients with more severe symptoms engage in recalibration for longer time intervals than those with milder symptoms²¹. Stroke is an example of a health condition where change in health is sudden, intense, and symptoms can persist over a long period of time; it is a health condition in which response shift may

occur. Stroke is a chronic disabling condition that can impact all aspect of function, perception, cognition, mood, speech, and quality of life ⁵ – a model for response shift.

The 1999/2000 age-standardized incidence of cerebral infarction in Canada has been estimated as 12.0/10,000; 14.4/10,000 for all types of acute stroke ³⁴. Rates of cerebral infarction have decreased in Quebec from 1988 – 2002, and increased for intracerebral hemorrhage ³⁵. It is estimated that approximately 300,000 Canadians are living with stroke ³⁶.

The functional problems encountered by persons with stroke are comprehensively classified by the WHO's International Classification of Function (ICF) ³⁷. All of the negative components of health (impairments, activity limitations and participation restrictions) are grouped under the umbrella term “*disability*”. All of the positive components (body structure and function, activity, and participation) are grouped under the umbrella term “*function*” ³⁷. See Table 1-1 for definitions of the terms. These terms are often used inconsistently and interchangeably without consideration that they define distinct concepts. Under the ICF model, an *impairment*, such as decreased strength or coordination of the limbs, causes limitations in *activities*, such as walking, which eventually could lead to *participation* restrictions (restricted ability to carry out usual activities in the community, such as work or volunteering). These participation restrictions and limitations in activity will in turn lead to further health impact through reduction of exercise capacity.

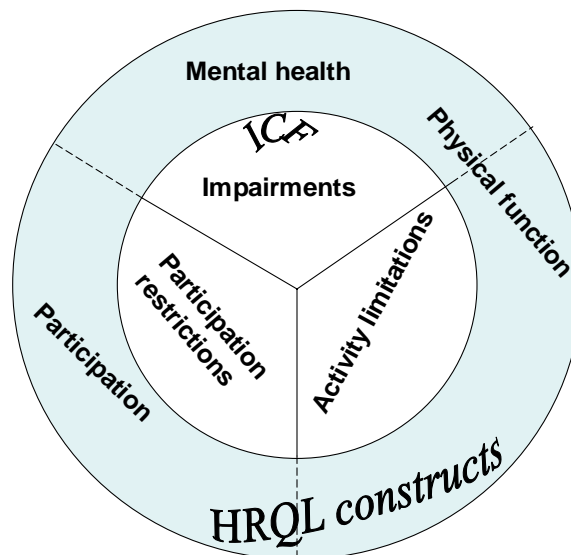
Table 1-1 The International Classification of Functioning (ICF) ³⁷

Functioning		Disability	
Body functions	physiological functions of body systems, including psychological functions	Impairment	problems in body function
Body structures	anatomical parts of the body, such as organs, limbs and their components	Impairment	problems in body structure
Activity	the execution of a task or action by an individual	Activity limitations	difficulties an individual may have in executing activities
Participation	involvement in a life situation, represents the societal perspective of functioning	Participation restrictions	problems an individual may experience in involvement in life situations
Contextual Factors			
Environmental factors	make up the physical, social and attitudinal environment in which people live and conduct their lives external to individuals and can have positive (facilitator) or negative (barrier) influence on the individual		
Personal factors	particular background of an individual's life and living situation and comprise features that are not part of health condition, e.g. gender, age, race, fitness, lifestyle, habits, social background, other health conditions		

The aspects of functioning that are impacted by stroke have been agreed upon by a consensus process conducted by the ICF Research Branch of the WHO Family of International Classifications Collaborating Center ³⁸. Examples of the content areas included in the ICF Core Set for stroke that are likely to have a direct impact on HRQL include (but are not exclusive to): energy functions, memory functions, emotional functions, exercise tolerance functions, communication, hand and arm use, walking, looking after one's health, community life, and recreation and leisure ³⁸.

As previously stated, physical, psychological, and social functioning are commonly described as constructs of HRQL¹⁹. Aspects that describe difficulty in physical functioning include impairments (such as decreased strength) or activity limitations (difficulty walking); difficulties with psychological functioning can include impairments in cognitive or emotional functioning; and difficulties in social functioning can be described as participation restrictions. See the Figure 1-1 for a demonstration of the association between the ICF domains and the constructs of HRQL.

Figure 1-1 Association between the ICF domains and the constructs of HRQL



Approximately one third of individuals with stroke have had very severe or severe strokes, while two-thirds are considered to have had mild or moderate strokes^{39, 40}. In general, those with large or severe strokes do less well than those with small or less

severe strokes ⁴¹. Twenty percent of people who have strokes die, 50% are discharged home, 10% are discharged to inpatient rehabilitation, and 15% of stroke survivors require long term care ⁵.

Moderate or severe activity limitations are still evident in 28% of stroke survivors after rehabilitation ⁴². Activity limitations and participation restrictions at 6-12 months post stroke have been noted in stroke survivors who were discharged from rehabilitation; improvements in activity that are achieved during rehabilitation may not be sustained at one year, with 33% of individuals worsening ^{43, 44}. The most commonly occurring activity limitations and participation restrictions include: having a meaningful activity during the day, doing household tasks, traveling, and activities of daily living ⁶. The Copenhagen Stroke Study found that activity limitations related to walking are common; only 64% of stroke survivors demonstrated independent indoor walking function when rehabilitation ended ³⁹.

HRQL measurement post stroke

HRQL has many applications to the study of individuals who have suffered a stroke. Increased quality of life is the ultimate goal in rehabilitation and other treatment post stroke ⁶. Treatment focuses on individual attributes / domains of HRQL which are important to the individual stroke client. It is therefore important to evaluate all aspects of stroke, which are reflected in HRQL domains, to be able to plan future services ⁶.

Renwick and Friefeld (1996) agree that enhancing HRQL is an overall goal in rehabilitation and should be important to rehabilitation theory, research, and practice. In practice, however, HRQL is not often measured in individuals undergoing rehabilitation,

but is assumed to have improved when other aspects, such as impairments and activity limitations improve ⁴⁵.

The study of QOL and HRQL post stroke can be used to inform treatment decisions, research, and health policy ^{9, 20, 46, 47}. HRQL is important to measure in stroke, as it incorporates the self perception of health status; neurological status and functional status don't provide enough information to be able to evaluate the impact of stroke on individuals who have survived a stroke ⁴⁸. Accurate information about HRQL as well as aspects of activity limitations is required to plan, provide, and allocate health services for stroke survivors ⁴⁹. Studies of cost-effectiveness of treatment, and studies of the effectiveness of acute, rehabilitation, and community stroke interventions often utilize HRQL outcomes ⁵⁰.

In addition to being useful in determining treatment effectiveness and health planning, it has been noted that HRQL as an outcome may be more important to stroke survivors than other outcomes such as impairments or activity limitations ⁵⁰. The best way to measure clinical improvements post stroke from a stroke survivor's point of view is with HRQL ⁵¹.

Some studies have shown lower HRQL in individuals with stroke compared to those without stroke; HRQL has also been shown to decrease over time ^{6, 48, 48, 52-55}. As well as the goal of improving HRQL through stroke rehabilitation, response shift itself is also a rehabilitation goal, though not typically stated as such by clinicians ^{6, 28}. Facilitating response shift is particularly important when impairments, activity limitations, and participation restrictions are not expected to recover fully, but improved HRQL is a goal.

Response shift post stroke

As mentioned, stroke can be used as a model for response shift due to the sudden onset and chronic nature of impairments, activity limitations, and participation restrictions that many individuals post stroke experience – a condition in which response shift may be expected.

The HRQL of survivors of stroke has been noted to be in the range of peers, despite decreased endurance and walking ability post stroke ⁵. Stroke survivors at six months post stroke have also demonstrated similar QOL compared to controls based on a 10 point QOL visual analogue scale; 6.8 and 7.8 respectively ⁶. A possible explanation for these findings is the presence of response shift.

Initial studies in the evaluation of response shift post stroke have both suggested and refuted the presence of response shift post stroke ⁵⁶. The methods utilized to determine the presence of response shift included: structural equation modeling (confirmatory factor analysis), the then test, and the Patient Generated Index (PGI) in three studies with stroke survivors in the first six months post stroke ^{4, 57, 58}.

The then test is a method that can be incorporated into the study design. A more detailed description and figure of the then test can be seen in the paper in Chapter two, “Much research needs to be done on Response Shift in order to accurately evaluate change in patient reported outcomes (PROs)”. In 2004, Ahmed et al used the then test with the 100 point visual analogue scale feeling thermometer of the Euroqol, a generic measure of HRQL, with participants post stroke to determine recalibration response shift between baseline and six weeks, and six and 24 weeks post stroke ⁴. Mean change in HRQL was

larger when measured retrospectively with the then test compared to conventional measurement of change (pre test – post test); the conventional measurement underestimated change in HRQL over time⁴. The then test is confounded to some extent by recall bias and results of the then test are also dependent on memory^{4, 21}. The memory of the participants with stroke was evaluated in this study; individuals with poor memory post stroke showed a greater variability in response shift compared to those shown with good memory⁴.

The PGI method suggested reconceptualization and reprioritization. The PGI is an individualized method, also described in the paper in Chapter two, “Much research needs to be done on Response Shift in order to accurately evaluate change in patient reported outcomes (PROs)”. Stroke survivors were evaluated with the PGI at six and 24 weeks post stroke⁵⁸. Ninety-two people completed evaluations at both time periods; 82 of those had changes in areas of importance and / or weights, which suggested the possibility of reconceptualization and/or reprioritization. Semi structured interviews were completed by 48 individuals to determine if response shift had occurred or if there were true changes in the domains / areas selected. Twenty-eight percent of those interviewed were determined to have experienced response shift⁵⁸.

Disadvantages of using the PGI to determine response shift in stroke survivors included the difficulty in administration, especially in the acute phase (6 weeks); some people were too confused and unable to follow instructions; or the evaluation was too time consuming⁵⁸. The PGI was also complex to interpret, requiring the semi-structured interview to determine response shift. The main advantage was that HRQL can be specifically measured in areas that are important to the individuals being assessed⁵⁸.

Structural equation modeling (SEM), using a measurement model based on the Short Form 36 (SF-36) in a cohort study of 238 clients post stroke and 392 controls did not identify response shift between one and six months post stroke⁵⁷. The SF-36 is a generic index of HRQL. Ahmed et al used the method described by Schmitt to determine if response shift was present in this study⁵⁹. A reconceptualization was suggested, however, early after stroke, possibly mediated by the stroke itself; the control group pre test was used as a proxy pre-stroke score for those with stroke⁵⁷.

SEM and Measurement Invariance

Structural equation modeling (SEM) is a multivariate technique used for the analysis of latent variables. Latent variables are variables which can not be measured directly or assessed reliably because of the presence of measurement error. Latent variables are associated with observed measured variables, thought to be reflective of the latent variables⁶⁰. These associations are referred to as paths or factor loadings. When unstandardized, these paths can be interpreted as regression equations. The measured variable is the outcome variable, the latent variable is the explanatory variable, and the factor loading is the slope; factor loadings in standardized solutions are interpreted as correlations between the measured and latent variable⁶¹. In a standardized solution, standardized parameter estimates are calculated from standardized variables which have a mean of zero and standard deviation that is equal to one. Unstandardized parameter estimates use variables that are in their original units⁶¹. The error variance (measurement error) of each relationship is also calculated. Relationships between latent variables or measurement errors in an unstandardized solution are covariances; in a standardized solution, correlations. To determine parameter estimates (paths, covariances,

error variances, and intercepts) in a model, a SEM computer program utilizes raw data files, covariance matrices, or correlation matrices with standard deviations; analyzing means will provide an intercept ⁶¹.

Once a model is developed, the fit of each model is assessed in numerous ways. The model chi-square (χ^2) is commonly reported in the SEM literature. As model χ^2 increases, the fit of a model become worse; the χ^2 tests the difference between the observed model and a model that has a perfect population fit ⁶¹. A non-significant χ^2 means that there is little difference between the models, suggesting a good fit. Models with large sample sizes, however, can often be rejected because the model χ^2 is affected by sample size ⁶². It is therefore suggested that a variety of fit indices be used. Models with a good fit to the data will demonstrate results that are uniform across various indices. Some researchers use a normed χ^2 ($\chi^2/\text{degrees of freedom}$) as a way to decrease the effect of the sample size on the χ^2 ; values up to 5.0 have been used to suggest a reasonable model fit ⁶¹. The root mean square error of approximation (RMSEA) measures the lack of fit in a model compared to the population. A value of ≤ 0.05 is considered a close fit, 0.05 to 0.08 is a reasonable fit and ≥ 0.10 a poor fit; the 90% confidence interval is also presented ⁶³. The comparative fit index (CFI) has a range of 0-1.0; a reasonably good fit is suggested with a CFI of greater than 0.90 ⁶⁴. Values of less than 0.10 are acceptable with the standardized root mean square residual (SRMR) ⁶¹.

Schmitt had described an SEM method in 1982 to identify reconceptualization and recalibration response shift ⁵⁹. In 2005, Oort proposed an alternate method to identify the presence of all three types of response shift with SEM ⁶⁵. The approach described by Oort utilizes confirmatory factor analysis (CFA) with longitudinal data to test for the presence

of response shift. All types of response shift (reconceptualization, reprioritization, and uniform and non-uniform recalibration) can be detected using this SEM approach. The method described by Schmitt does not identify reprioritization or non-uniform recalibration; it interprets changes in factor loadings differently from the approach described by Oort. In the approach described by Schmitt, constraints are added to the model to identify response shift; in the approach described by Oort, constraints are removed to identify response shift^{59, 65, 66}. The constraints added or removed are equality constraints, where a parameter estimate at one time is made to be equal to another time⁶¹.

In the approach described by Oort, changes in the factor loading patterns over time suggest reconceptualization response shift⁶⁵. This is demonstrated with a measured variable loading onto one latent variable, and at a later point in time loading onto another latent variable. Changes in the values of factor loadings represents reprioritization⁶⁵; the factor loading of a measured variable to a latent variable may become stronger or weaker over time, reflecting changes in values or priorities. Two types of recalibration have been described: uniform recalibration occurs when the whole measurement scale is recalibrated, while non-uniform recalibration occurs when only a portion of the measurement scale is recalibrated⁶⁵. Uniform recalibration is determined by differences between intercepts across occasions and non-uniform recalibration by differences between error variances across occasions⁶⁵.

The evaluation of response shift with SEM involves the evaluation of measurement invariance. Measurement invariance occurs when parameter estimates across times do not change. If parameter estimates do vary (e.g. factor loading patterns, factor loading values, intercepts, or error variances), reconceptualization, reprioritization, uniform recalibration,

or non-uniform recalibration response shifts are thought to have occurred. Various degrees of measurement invariance have been described⁶⁷. Weak invariance refers to invariance of factor loadings across models or over time⁶⁷; it is necessary to determine that populations are similar, that the latent variables in two models have a constant meaning⁶⁸. Strong invariance refers to the factor loadings as well as the intercepts of measured variables being invariant, necessary for inferring true change of the latent variables in a population^{67, 68}. Strict invariance refers invariance of factor loadings, intercepts, and error variances across models, leading to a comparison of the model across populations that is unbiased^{67, 68}. In a discussion of two methods used to identify response shift with SEM^{57, 65}, Donaldson questioned the interpretation of any lack of invariance as identifying a specific type of response shift⁶⁸. Factor loading invariance is considered first in classical factor analysis; if the unstandardized regression coefficients are the same across populations or time, it is assumed that the models are the same⁶⁸. Factor loading invariance applies to both reconceptualization (change in factor loading pattern) and reprioritization (change in factor loading values). A hierarchy of assessment of response shift has been suggested: first rule out gamma change (reconceptualization), then rule out beta change (recalibration); you are able to identify alpha change (real change)²⁵.

Implications of Response Shift Measurement - Why is it important to study

Response Shift?

Measurement of change in HRQL will be inaccurate if response shift is present due to different conceptualization / prioritization / or calibration over time from pre to post measurement. Response shift must therefore be taken into account. It has been

demonstrated that change in HRQL can be underestimated when recalibration occurs, as measured by the then test ⁴. The effect size of HRQL change can be estimated, taking response shift into account, after evaluation with SEM: HRQL change is generally larger when response shift is accounted for ⁶⁹.

Increased knowledge of response shift will therefore affect the way in which HRQL measures are used in clinical, research, and policy decisions. It has been stated that the development of an estimation of HRQL which takes into account recalibration, reprioritization, and reconceptualization is crucial for being able to interpret the results of studies of interventions post stroke, such as studies of health services delivery ⁵⁶.

Further study of response shift will also add to the understanding of how people adapt to health with chronic illness, such as stroke ²¹. In this vein, response shift is often the focus of medical treatments, including rehabilitation ²⁸. For example, a health professional may want an individual to have a ‘good outlook’ (good HRQL) despite limitations in physical function, with a focus on the positive. Rehabilitation interventions with individuals may teach response shift, therefore response shift assessment should be part of the self report measures used ⁸. In a self-management program, the goal is often to induce response shift. The Health Education Impact Questionnaire was developed to identify response shift among individuals who participated in a self-management course for arthritis ⁷⁰. In this case, the goal is response shift, and response shift is specifically measured after the intervention. Further study into response shift will lead to more information in this new area where more needs to be discovered. The following chapter is a review of response shift research with suggested future research priorities.

Reference List

- (1) FDA. Guidance for Industry: Patient Reported Outcomes: Use in Medical Product Development to Support Labeling Claims: Draft Guidance.FDA, Rockville,MD. 2006 Jan 2.
- (2) Shimozuma K, Imai H, Ohsumi S, Ono M. Recent Topics in Health Outcomes Research in Oncology. *Breast Cancer* 2007;14(1):60-5.
- (3) Sprangers MA, Schwartz CE. Integrating Response Shift into Health-Related Quality of Life Research: A Theoretical Model. *Social Science and Medicine* 1999;48:1507-15.
- (4) Ahmed S, Mayo N, Wood-Dauphinee S, Hanley J, Cohen S. Response Shift Influenced Estimates of Change in Health-Related Quality of Life Poststroke. *Journal of clinical epidemiology* 2004;57:561-70.
- (5) Mayo N, Wood-Dauphinee S, Ahmed S, Gordon C, Higgins J, McEwen S, Salbach N. Disablement Following Stroke. *Disability and Rehabilitation* 1999;21(5/6):258-68.
- (6) Mayo N, Wood-Dauphinee S, Cote R, Durcan L, Carlton J. Activity, Participation, and Quality of Life 6 Months Poststroke. *Archives of Physical Medicine and Rehabilitation* 2002;83:1035-42.
- (7) Anderson C, Carter K, Brownlee W, Hackett M, Broad J, Bonita R. Very Long-Term Outcome After Stroke in Auckland, New Zealand. *Stroke* 2004;35:1920-4.

- (8) Schwartz C, Andresen EM, Nosek M, Krahn G, RRTC Expert Panel on Health Status Measurement. Response Shift Theory: Important Implications for Measuring Quality of Life in People With Disability. *Archives of Physical Medicine and Rehabilitation* 2007;88:529-36.
- (9) Pipitone N, Scott D, Vincent K. Measurement of Health-Related Quality of Life in Rheumatology: Current Concepts and Controversies. *International Journal of Health Promotion and Education* 2002;40(4):105-10.
- (10) White C, Lauzon S, Yaffe MJ, Wood-Dauphinee S. Toward a Model of Quality of Life for Family Caregivers of Stroke Survivors. *Quality of Life Research* 2004;13:625-38.
- (11) Acquadro C, Chassany O, Longin J, Patrick D. Educational Program on Health-Related Quality of Life (HRQL) in Clinical Trials - workbook. Cochrane Health-Related Quality of Life Methods Group, MAPI Research Trust, the ERIQA Group; 2004.
- (12) Bowling A. *Measuring Disease - A Review of Disease-Specific Quality of Life Measurement Scales*. Buckingham: Open University Press; 1995.
- (13) McDowell I, Newell C. General Health Status and Quality of Life. *Measuring Health: A Guide to Rating Scales and Questionnaires*. 2nd ed. Oxford: Oxford University Press; 1996. p. 380-492.
- (14) Padilla G, Frank-Stromberg M, Koresawa S. Single Instruments for Measuring Quality of Life. In: Frank-Stromberg M, Olsen S, editors. *Instruments for*

Clinical Health-Care Research. 3rd ed. Boston: Jones and Bartlett Publishers International; 2004. p. 128-49.

- (15) Flanagan J. A Research Approach to Improving our Quality of Life. *Am Psychol* 1978;33:138-47.
- (16) Flanagan J. Measurement of Quality of Life: Current State of the Art. *Archives of Physical Medicine & Rehabilitation* 1982;63(2):56-9.
- (17) WHO. Introducing the WHOQOL Instruments. *WHO website* 2004; Available at: URL: <http://www.who.int/evidence/assessment-instruments/qol/q11.htm>.
- (18) WHO. Structure of the WHOQOL-BREF. *WHO website* 2004; Available at: URL: <http://www.who.int/evidence/assessment-instruments/qol/q16.htm>.
- (19) Sprangers MA. Quality-of-life assessment in oncology. Achievements and challenges. *Acta oncologica Stockholm, Sweden* 2002;41(3):229-37.
- (20) Torrance G. Utility Approach to Measuring Health-Related Quality of Life. *Journal of Chronic Disease* 1987;40(6):593-600.
- (21) Schwartz CE, Sprangers MAG, Carey A, Reed G. Exploring response shift in longitudinal data. *Psychology and Health* 2004 February; Vol 19(1):51-69.
- (22) Patrick D, Erickson P. *Health Status and Health Policy - Quality of Life in Health Care Evaluation and Resource Allocation*. Oxford: Oxford University Press; 1993.

- (23) Norman G. Hi! How are you? Response shift, implicit theories and differing epistemologies. *Quality of life research an international journal of quality of life aspects of treatment, care and rehabilitation* 2003 May;12(3):239-49.
- (24) Wilson I, Cleary P. Linking Clinical Variables with Health-Related Quality of Life. A Conceptual Model of Patient Outcomes. *JAMA* 1995;273:59-65.
- (25) Schwartz CE, Sprangers MA. Methodological Approaches for Assessing Response Shift in Longitudinal Health-Related Quality of Life Research. *Social Science and Medicine* 1999;48:1531-48.
- (26) Razmjou H, Yee A, Ford M, Finkelstein J. Response Shift in Outcome Assessment in Patients Undergoing Total Knee Arthroplasty. *The Journal of Bone and Joint Surgery* 2006;88-A(12):2590-5.
- (27) Ring L, Hofer S, Heuston F, Harris D, O'Boyle C. Response Shift Masks the Treatment Impact on Patients Reported Outcomes (PROs): the Example of Individual Quality of Life in Edentulous Patients. *Health and Quality of Life Outcomes* 2005 September 7;3(55).
- (28) Wilson IB. Clinical Understanding and Clinical Implications of Response Shift. *Social Science and Medicine* 1999;48:1577-88.
- (29) Schwartz C, Sprangers M. *Adaptation to Changing Health - Response Shift in Quality-of-Life Research*. Washington: American Psychological Association; 2000.

- (30) Golembiewski R, Billingsley K, Yeager S. Measuring Change and Persistence in Human Affairs: Types of Change Generated by OD Designs. *Journal of Applied Behavioural Sciences* 1976;12(2):133-57.
- (31) Armenakis A. A Review of Research on the Change Typology. In: Pasmore W, Woodman R, editors. *Research in Organizational Change and Development*. 2 ed. Greenwich: JAI Press Inc.; 1988. p. 163-94.
- (32) Schwartz C, Sprangers M. Discussion: Theoretical Reflections. In: Schwartz C, Sprangers M, editors. *Adaptation to Changing Health - Response Shift in Quality-of-Life Research*. Washington: American Psychological Association; 2000. p. 73-7.
- (33) Brossart D, Clay D, Willson V. Methodological and Statistical Considerations for Threats to Internal Validity in Pediatric Outcome Data: Response Shift in Self-Report Outcomes. *Journal of pediatric psychology* 2002;27(1):97-107.
- (34) Johansen H, Wielgosz A, Nguyen K, Fry RN. Incidence, Comorbidity, Case Fatality and Readmission of Hospitalized Stroke Patients in Canada. *Canadian Journal of Cardiology* 2006;22(1):65-71.
- (35) Mayo N, Nadeau L, Daskalopoulou S, Côté R. The Evolution of Stroke in Quebec. *Neurology* 2007;2007(68):1122-7.
- (36) Heart and Stroke Foundation of Canada. General Info - Stroke Statistics. *Heart and Stroke Foundation of Canada website* 2002 September 3; Available at: URL: www2.heartandstroke.ca. Accessed April 11, 2008.

- (37) World Health Organization. *International Classification of Functioning, Disability and Health*. Geneva: World Health Organization; 2001.
- (38) Geyh S, Schouten J, Dickson H, Frommelt P, Omar Z, Kostanjsek N, Ring H, Stucki G. ICF Core Sets for Stroke. *Journal of Rehabilitation Medicine* 2004;suppl 44:135-41.
- (39) Jorgensen H, Nakayama H, Raaschou H, Vive-Larsen J, Stoier M, Olsen T. Outcome and Time Course of Recovery in Stroke. Part 1: Outcome. The Copenhagen Stroke Study. *Archives of Physical Medicine and Rehabilitation* 1995;76:399-405.
- (40) Mayo N. Epidemiology and Recovery of Stroke. *Physical Medicine and Rehabilitation: State of the Art Reviews* 1998;12(3):355-65.
- (41) Jorgensen H, Reith J, Nakayama H, Kammergaard L, Raaschou H, Olsen T. What Determines Good Recovery in Patients With the Most Severe Strokes? The Copenhagen Study. *Stroke* 1999;30:2008-12.
- (42) Jorgensen H, Nakayama H, Pedersen P, Kammergaard L, Raaschou H, Olsen T. Epidemiology of Stroke-Related Disability. *Clinics in Geriatric Medicine* 1999;15(4):785-99.
- (43) Corr S, Bayer A. Poor Functional Status of Stroke Patients After Hospital Discharge: Scope for Intervention? *British Journal of Occupational Therapy* 1992;55(10):383-5.

- (44) Paolucci S, Grasso M, Antonucci G, Troisi E, Morelli D, Coiro P, Bragoni M. One-Year Follow-Up in Stroke Patients Discharged from Rehabilitation Hospital. *Cerebrovascular Disease* 2000;10(1):25-32.
- (45) Renwick R, Friefeld S. Quality of Life and Rehabilitation. In: Renwick R, Brown I, Nagler M, editors. *Quality of Life in Health Promotion and Rehabilitation*. Thousand Oaks: SAGE Publications; 1996. p. 26-37.
- (46) Guyatt G, Feeny D, Patrick D. Measuring Health-Related Quality of Life. *Annals of Internal Medicine* 1993;118:622-9.
- (47) WHO. The Uses of the WHOQOL Instruments. *WHO website* 2004; Available at: URL: <http://www.who.int/evidence/assessment-instruments/qol/ql4.htm>.
- (48) Suenkeler I, Nowak M, Misselwitz B, Kugler C, Schreiber W, Oertel W, Back T. Time course of Health -Related Quality of Life as Determined 3, 6 and 12 Months After Stroke - Relationship to Neurological Deficit, Disability and Depression. *Journal of Neurology* 2002;249:1160-7.
- (49) Hackett M, Duncan J, Anderson C, Broad J, Bonita R. Health-Related Quality of Life Among Long-Term Survivors of Stroke - Results from the Auckland Stroke Study, 1991-1992. *Stroke* 2000;31:440-7.
- (50) Sturm J, Osborne R, Dewey H, Donnan G, Macdonell A, Thrift A. Brief Comprehensive Quality of Life Assessment After Stroke - The Assessment of Quality of Life Instrument in the North East Melbourne Stroke Incidence Study (NEMESIS). *Stroke* 2002;33:2888-94.

- (51) Poissant L, Mayo N, Wood-Dauphinee S, Clarke A. The Development and Preliminary Validation of a Preference-Based Stroke Index (PBSI). *Health and Quality of Life Outcomes* 2003 September 10;1(43).
- (52) Grootendorst P, Feeny D, Furlong W. Health Utilities Index Mark 3: Evidence of Construct Validity for Stroke and Arthritis in a Population Health Survey. *Medical Care* 2000;38(3):290-9.
- (53) Duncan P, Weinberger M, Goldstein L, Bonito A, Witter D, Enarson C, Matchar D. Health Status of Individuals with Mild Stroke. *Stroke* 1997;28(4):740-5.
- (54) Li L, Young D, Xiao S, Zhou X, Zhou L. Psychometric Properties of the WHO Quality of Life Questionnaire (WHOQOL-100) in Patients with Chronic Diseases and Their Caregivers in China. *Bulletin of the World Health Organization* 2004;82(7):493-502.
- (55) Hopman W, Verner J. Quality of Life During and After Inpatient Stroke Rehabilitation. *Stroke* 2003;34:801-5.
- (56) Ahmed S, Mayo N, Wood-Dauphinee S, Hanley J, Cohen S. The Structural Equation Modeling Technique Did Not Show a Response Shift, Contrary to the Results of the Then Test and the Individualized Approaches. *Journal of clinical epidemiology* 2005;58:1125-33.

- (57) Ahmed S, Mayo N, Corbiere M, Wood-Dauphinee S, Hanley J, Cohen R. Change in Quality of Life of People with Stroke Over Time: True Change or Response Shift? *Quality of Life Research* 2005;14:611-27.
- (58) Ahmed S, Mayo N, Corbiere M, Wood-Dauphinee S, Hanley J, Cohen R. Using the Patient Generated Index to Evaluate Response Shift Post Stroke. *Quality of Life Research* 2005;14:2247-57.
- (59) Schmitt N. The Use of Analysis of Covariance Structures to Assess Beta and Gamma Change. *Multivariate Behavioural Research* 1982;17:343-58.
- (60) Hoyle R. *Structural Equation modelling - Concepts, Ideas, and Applications*. Thousand Oaks: SAGE Publications; 1995.
- (61) Kline R. *Principles and Practice of Structural Equation Modeling*. 2 ed. New York: The Guilford Press; 2005.
- (62) Hays R, Revicki D, Coyne K. Application of Structural Equation Modeling to Health Outcomes Research. *Evaluation and the Health Professions* 2005;28(3):295-309.
- (63) Browne M, Cudeck R. Alternative Ways of Assessing Model Fit. In: Bollen K, Long J, editors. *Testing Structural Equation Models*. London: Sage Publications; 1993. p. 136-62.
- (64) Jaccard J, Wan C. *LISREL Approaches to Interaction Effects in Multiple Regression*. Thousand Oaks: Sage Publications; 1996.

- (65) Oort FJ. Using Structural Equation Modeling to Detect Response Shifts and True Change. *Quality of Life Research* 2005;14:587-98.
- (66) Borsboom G, Korfage IJ, Essink-Bot ML, Duivenvoorden H. Letter to the Editor - The Structural Equation Modeling Technique Did Not Show a Response Shift, Contrary to the Results of the Then Test and the Individualized Approaches. *Journal of clinical epidemiology* 2007;60:426-8.
- (67) Meredith W, Teresi J. An Essay on Measurement and Factorial Invariance. *Medical Care* 2006;44(11 Suppl 3):S69-S77.
- (68) Donaldson G. Structural Equation Models for Quality of Life Response Shifts: Promises and Pitfalls. *Quality of Life Research* 2005;14:2345-51.
- (69) Oort FJ, Visser M, Sprangers MA. An Application of Structural Equation Modeling to Detect Response Shifts and True Change in Quality of Life Data from Cancer Patients Undergoing Invasive Surgery. *Quality of Life Research* 2005;14:599-609.
- (70) Osborne R, Hawkins M, Sprangers M. Change of Perspective: A Measureable and Desired Outcome of Chronic Disease Self-Management Intervention Programs That Violates the Premise of Preintervention / Postintervention Assessment. *Arthritis and Rheumatism* 2006;55(3):458-65.

Chapter 2: Much research needs to be done on Response Shift in order to accurately evaluate change in patient reported outcomes (PROs)

Submitted to 'Quality of Life Research' and undergoing revisions as of August 19, 2008

Ruth Barclay-Goddard MHSc, BMR(PT) ¹

Joshua Epstein MA²

Nancy E. Mayo, BSc(PT), MSc, PhD³

1. Department of Physical Therapy, School of Medical Rehabilitation and Department of Community Health Sciences, University of Manitoba, R106-771 McDermot Ave. Winnipeg, Manitoba, Canada R3E 0T6

email barclayg@cc.umanitoba.ca Phone (204)977-5637, fax (204)789-3927

2. Department of Pharmaceutical Economics and Policy, School of Pharmacy, University of Southern California, 1540 East Alcazar St, CHP-140, Los Angeles, California, USA 90089-9004 email jdepstei@usc.edu Phone (323) 442-1825, fax (323) 442-1462

3. Department of Medicine and School of Physical and Occupational Therapy, McGill University; Division of Clinical Epidemiology; Division of Geriatrics, McGill University Health Center, Royal Victoria Hospital Site, R4.29, 687 Pine Ave W, Montreal, QC, H3A 1A1

email nancy.mayo@mcgill.ca Phone (514) 934 1934 36922, fax (514) 843 1493

Much research needs to be done on Response Shift in order to accurately evaluate change in patient reported outcomes (PROs)

Abstract

Aim The objective of this paper is to review current methodologies of response shift research in patient reported outcomes to facilitate and stimulate further research in this area. **Methods** This paper is a narrative review of research in response shift. **Results** The following research priorities emerged : 1) Improve the theoretical underpinnings of response shift (e.g. come to a clear definition of response shift, empirically test theoretical models and develop new models of response shift if needed) 2) Continue to evaluate the validity of current and emerging methods (e.g. further evaluation of the then test and SEM methods, and concurrent validity between current and emerging methods) 3) Determine the clinical relevance of response shifts 4) Utilize guidelines for the reporting of response shift. **Conclusions** With the adoption of these research priorities, we anticipate that the theories and processes of response shift will be better understood, current methods to analyze this phenomenon will be improved while new ones may also be developed, and the clinical relevance and impact of response shift in measuring changes in HRQL will be shown.

Key words:

Quality of life

Response shift

Patient reported outcomes

Abbreviations:

PRO's	patient reported outcomes
FDA	Food and Drug Administration
HRQL	health related quality of life
PGI	Patient Generated Index
SEIQOL	Schedule for the Evaluation of Individual Quality of Life
HEI-Q	Health Education Impact Questionnaire
CFA	confirmatory factor analysis
SEM	structural equation modeling
EQ5D	Euroqol – five dimensions
DIF	differential item functioning
MID	minimal important difference
VAS	visual analogue scale
ISOQOL	International Society for Quality of Life Research

Introduction

There is an increased acknowledgement of the importance of patient reported outcomes (PROs) in assessing the impact of interventions including drugs, as evidenced by the recent United States Food and Drug Administration (FDA) draft guidance and European guidance for these measures [1,2]. As PROs become more established, there is increased recognition and concern that a phenomenon known as response shift will need to be considered when interpreting PRO results.

When using PROs such as measures of health related quality of life (HRQL), values may be incorrectly interpreted if response shift is not taken into account [3,4,5]. Response shift, which is a change in the meaning of self-evaluation, is evident in individuals with chronic illnesses or disability. Many of these individuals report HRQL similar to that of healthy individuals, despite ongoing activity limitations [6,7,8]. In some instances, response shift is a desired outcome, such as in palliative care; in other instances, response shift can be a confounding factor in evaluating treatment outcome [4,9]. The presence of response shift also has potential implications for research involving HRQL outcomes as well as for the clinical interpretation of change in HRQL over time.

Pioneers and current leaders in response shift theory and research, Carolyn Schwartz and Mirjam Sprangers, have defined response shift as “a change in the meaning of one’s self-evaluation of a target construct as a result of: (a) a change in the respondent’s internal standards of measurement (i.e. scale recalibration); (b) a change in the respondent’s values (i.e. the importance of component domains constituting the target construct) or (c) a redefinition of the target construct (i.e. reconceptualization)” [10]. The authors refer

the reader to two landmark books that address response-shift and have served as the foundation for others who have joined this research field [11,12].

The aim of this review is to present in one document sufficient detail about the theory, study design, and statistical approaches available used to identify and quantify response shift in order to stimulate research in this area. Building on earlier reviews [10,11,12,13] this paper presents strengths and weakness of the different approaches and indicates areas requiring future research.

What is Response Shift?

Carolyn Schwartz has equated response shift to a homeostatic process designed to keep a person's feeling of well-being within a narrow positive range despite considerable health adversity[14]. This has been associated to the autonomic control of blood pressure and temperature and it may even be genetically determined [14]. Geoff Norman in his thoughtful paper, "Hi, How Are You?" tells the story of his father-in-law with terminal lung cancer saying each day as he deteriorated that he was "Oh, pretty good, considering"[15] and describes this as an example of response shift.

Response shift may also be a planned or anticipated effect of interventions such as self management and psycho-social programs, rehabilitation, and palliative care [8,9,16,17,18,19]. Facilitating response shift may be especially important in situations where symptoms and functions may not improve dramatically but better quality of life is desired, notably palliative care and management of certain chronic health states such as stroke or arthritis.

Response shift research is grounded in the fields of educational training interventions [20] and organizational change [21]. However, response shift can occur in any field where self-reported measures are evaluated over time. The working definition of response shift with respect to HRQL is a change in the meaning of an individual's self-evaluation of their HRQL as a result of a change in their internal standards, values, and/or concepts of HRQL [13]. These changes are also described as scale recalibration, reprioritization, and reconceptualization, respectively.

As an example of recalibration, consider the scoring of pain on a 10 point pain scale, where zero is no pain, and 10 is the worse pain imaginable. A person may rate a severely bruised knee as 8/10 on this pain scale, then later acquire a kidney stone, the passing of which was much more painful, rated close to 10. She then realizes that the earlier knee pain was probably only a 4/10, reflecting recalibration.

With reprioritization, an individual values the HRQL concepts of physical activity, socialization, and family as important. After a health scare, this person may find the same concepts important, but the order of importance (the value) changes to family first, socialization, and lastly, physical function.

With regards to reconceptualization, a healthy woman feels that the most important concept or aspect of HRQL is her high energy level, physical function, and participation in work related roles. This person experiences a major health event; she now feels that family, pain free moments and walking out of doors are the most important part of her HRQL. With intervention, this person is now able to do more of her work-related

activities but has not really changed in symptoms or function. HRQL over time has undergone reconceptualization.

In all of these examples, the concept of HRQL is not the same as before and cannot be compared over time due to changes in internal standards, values, and/or concepts. HRQL may have changed, but it will not be captured using traditional pre and post measures because of these changes.

If a response shift has occurred, calculations of change over time will yield values that are difficult to interpret, because of changes in internal standards [3,9]. Some people will have made true change in their HRQL and others will have undergone response shift potentially masking the effects of the variable under study, be it an intervention or the passage of time.

Theoretical Models of Response Shift

In 1999 Sprangers and Schwartz proposed a theoretical model of response shift [13], in which a change in health acts as a catalyst invoking behavioral and cognitive processes. These processes in turn are the mechanisms producing a change in internal standards, values or conceptualization, which then affects perceived quality of life [13]. Initial empirical testing of this model indicates that there is evidence for a “catalyst – to – mechanism – to – response shift” path but less evidence for a path between these components and perceived quality of life [22].

In 2004 Rapkin and Schwartz offered an updated model of response shift which built on the 1999 model by adding frame of reference (termed HRQL appraisal) as an important component of HRQL assessment [23]. For example, to judge her current health status, a

person may use as a frame of reference, her prior health status, an ideal health status, or the presumed health status of someone else with the same condition [8,23]. The Quality of Life Appraisal Profile was introduced by Rapkin and Schwartz [23], to measure appraisal processes used by an individual.

People can change their frame of reference and comparisons over time or with different health and treatment experiences [24] making it essential that this information is elicited each time a HRQL rating is obtained. Different aspects of HRQL may also have different frames of reference. A person may use a prior time point as a frame of reference for pain, but may use the “person in the next bed” as a frame of reference for function. For example, individuals undergoing palliative chemotherapy appear to maintain their quality of life even when physical health deteriorates, when they consider that they are better off than another person in a similar situation [25].

Social comparisons that invoke a positive frame (“I could be as good as ...”; or “I am not so badly off as ..”) have been found to be associated with better quality of life adjustment, while negatively interpreted comparisons may be associated with worse adjustment [26]. Further development of tools to identify and measure frames of reference and to apply these tools when eliciting quality of life ratings is an important area of research. This would expand our understanding of response shift and would contribute to the further development of its theoretical basis.

Study Design to Address Response Shift

Knowing that many health situations and interventions can result in response shift, a researcher has a number of methodological approaches available when designing a study where there is the potential for response shift. These can be grouped under individualized methods, retrospective ratings or direct questioning, all of which are considered time consuming and burdensome for the individuals involved [10,27].

Retrospective Ratings

Retrospective ratings of past health states using the ‘then test’ is the most commonly used method for measuring response shift. Researchers have incorporated the then test with a variety of PRO instruments and disease states. With traditional study design, changes in PROs are calculated by taking the difference between assessments before (pretest) and after (posttest). The then test is a retrospective re-assessment of the baseline (pretest) and is assessed at the same time as the post-test. As both evaluations are done at the same time point, it is assumed that the same standards, values, and concepts of HRQL are used for both assessments. The then test has been used successfully to study a variety of patient populations [3,28,29,30].

The difference between the then test and the pre-test is considered the response shift effect. The difference between the post-test and then test is used as the measure of response shift adjusted change. See Figure 2-1. The then test is assumed to measure recalibration [10].

There are drawbacks to the then test. The effect of recall ability on these retrospective evaluations requires further scrutiny. Individuals with poor memory post stroke have

shown a greater variability in response shift compared to those shown to have good memory [3]. Measuring recall ability in this context is necessary to understand the reliability and validity of the then test as a means to measure prior HRQL.

Geoff Norman offered criticism regarding the validity of the then test based on the heuristics involved with providing a renewed judgment on a respondent's prior health [15]. Norman reasoned that if the respondent considers her current health and then internally calculates what her prior health was based on how much she thinks she has changed, the then test is inherently biased because she has not directly recalled her prior health state. Norman argues that in this case the first rating is unbiased and the then test rating would be flawed. However, if in reality, the person recalls prior health accurately and gives a renewed judgment based on what she knows now, the then test would be more valid than the Pretest. These heuristics need to be examined further in relation to the then test and response shift.

Individualized Methods

Using an individualized method, an individual is asked to select and rate the value of different domains of quality of life [10]. Two individualized methods that have been used in the evaluation of response shift are the Patient Generated Index (PGI) [27] and the Schedule for the Evaluation of Individual Quality of Life (SEIQOL) [5].

In the PGI, a person is asked to choose five areas of their life, rate their ability in these areas, and then dispense 12 tokens across these areas to reflect their importance [27]. This is shown in Figure 2-2. Over time, reprioritization is reflected by a change in the order of the domains chosen and reconceptualization is reflected by a change in the content of the

domains [10]. In a study of stroke survivors, use of the PGI suggested that persons experienced reconceptualization and reprioritization of HRQL over time [27]. For example, one stroke survivor selected the areas of handwriting and driving at 6 weeks post stroke and gave one token to handwriting and 10 tokens to driving (one token went to a non-health related area). At 24 weeks, the individual chose the same areas, but gave five tokens to handwriting and seven to driving. This demonstrates reprioritization, as the areas stayed the same, but values for each area changed [27].

The SEIQOL has also been used to identify response shift in HRQL. O'Boyle et al [31] give the example of an individual who was terminally ill. In the months prior to his death, the areas important to him changed (reconceptualization), as well as the importance of each area (reprioritization).

The advantage of these procedures is that an individual can define those areas of HRQL of personal importance. Although these methods can identify that a response shift occurred in a particular person, these approaches are not easily converted into a numerical value of the response shift effect. Further research using these methods will help advance our understanding of when response shift occurs and what characteristics are associated with patients who do and do not undergo such changes.

Direct Questioning

To enable people to describe their experiences and potentially evaluate all aspects of response shift, individuals may be directly questioned about their HRQL using semi structured (qualitative) interviews [10,27]. Questions can be directed to a specific aspect of response shift and the commentary can assist in future hypothesis generation.

This method is time consuming for both data collection and analysis. It also relies on cognitive awareness of response shift, which is not always present [10]. Among the 92 stroke survivors who completed the PGI in the study by Ahmed et al [27], only 47 were able to provide any information related to their responses through these qualitative interviews, demonstrating that in many clinical populations this type of interviewing may not be feasible.

Vignettes Ratings

Rating vignettes before and after an intervention that is hypothesized to induce a response shift is a newer method of identifying people who have undergone a response shift. Significant changes in rating scores assigned to a vignette identify response shift. This method has been successfully incorporated into a prostate cancer setting using three vignettes for hypothetical health states representing different side-effects of prostate cancer treatment [32]. Vignettes describing side effects of prostate cancer treatment (e.g. urinary leakage, bowel cramps, erectile dysfunction) were rated as less detrimental at one month after diagnosis as compared to two months prior to diagnosis. These results reflect reprioritization response shift [32].

Questionnaire

The Health Education Impact Questionnaire (HEI-Q) was developed to identify response shift among individuals who participated in a self-management course for arthritis. The nine item questionnaire uses a 7-point scale to indicate the degree of response shift which occurs [9]. In question four, the anchor at one end states “I now realize that before the course I did fewer healthy activities than I thought I did”, while the opposite anchor is “I

now realize that before the course I did more healthy activities than I thought I did”. The middle point is “the same” [9]. Item scaling therefore reflects negative, positive, or absent response shift. Response shift was described as negative when people realize that they were worse at a prior point, positive when they feel that they were better at a previous point, or absent when the two ratings were the same [9]. If negative response shift is not accounted for, a treatment effect may be underestimated, while positive response shift may overestimate a treatment effect [9].

Of the 121 respondents in this study, 87% showed recalibration response shift on some part of the questionnaire after participation in the self-management course. This study provides an example where the goal of intervention is response shift.

Statistical Methods to Address Response Shift

Statistical methods are often used where it is impractical to incorporate response shift evaluation into the study design or when doing secondary data analysis. Statistical analysis combines data across individuals. The major disadvantage of these group level techniques is that individual effects may be masked when observing data at the group level.

Factor analysis

The analysis of covariance structures approach using confirmatory factor analysis (CFA) has been used to identify both reconceptualization and recalibration by evaluating changes in factor structure and factor loading over time, respectively [33,34]. Ahmed et al used the CFA method described by Schmitt in a post-stroke population [33,34].

Response shift was not identified between one and six months despite it being suggested

using individualized methods [34], as there were no differences in the factor loadings over time.

Oort described an alternate method of determining if a response shift has occurred with structural equation modeling (SEM), again using CFA [35]. In SEM, the relationships (such as factor loading) between latent variables and measured variables are modeled. Latent variables represent constructs that can not be measured directly and consist of the commonality between several measured variables of closely related constructs. For example, HRQL cannot be measured directly and so in the SF-36, two latent variables to represent physical and mental health are constructed from the 8 sub-scales using the factor weights. In Oort's framework, reconceptualization may be suggested by changes in the factor loading patterns, reprioritization by the change in the values of factor loading, and uniform recalibration by differences between intercepts across occasions. See Figure 2-3. This method has been demonstrated in those with cancer undergoing invasive surgery, with patients having bone metastases, with participants in health education programs, and in stroke survivors [36,37,38,39]. It has been suggested that a substantial number of individuals in a study need to experience response shift for it to appear in an SEM model, as response shift is averaged across the group [34,35].

Oort has presented a definition of response shift as a 'special case of measurement bias' [40]. Measurement bias is a 'violation of measurement invariance'; time is considered a possible violator of measurement invariance. For example, in SEM, measurement invariance may be seen as no change of factor loadings over time (measurement does not vary); a violation of measurement invariance (response shift) is demonstrated in part as changes of factor loadings over time [40].

Growth curve analysis

Growth curve analysis has been described as a method for inferring response shift [41]. Growth curves provide an estimate of change over time for observed behaviours or functions related to HRQL and individuals' perception of these constructs. Response shift is inferred by comparing the form of the growth curve between measured and perceived domains of HRQL. A response shift may have taken place if there is disengagement between the two curves [10]. Growth curve modeling has been used with structural equation modeling to model the timing and shape of response shift [42]. A variety of shapes of response shift could occur depending on the time between measurements, the variables measured, and the type of participants [42]. The study of growth curve analysis in combination with other methods is an area for further research.

Residual Analysis

Not dissimilar to the growth curve analysis which attempts to capture the discordance between observed behaviours or functions and the person's perception of these constructs is a new method proposed by Mayo which uses the residuals generated from a growth model where measured health behaviours or functions are used to predict perceived health. Statistical methods typically give information at the group level of analysis and up to now only design imposed methods have been able to yield information at the individual level. However, residual analysis is at an individual level as each person in the analysis is assigned a residual at each time point depending on how far away from their predicted HRQL is their reported HRQL [43]. Mayo hypothesized that large fluctuations in differences between reported and predicted health-related quality of life (HRQL) could

be used to identify the presence of response shift and classify persons into groups based on the direction and timing of the response shift [43].

Using the experience of stroke survivors during the first year post-stroke, reported HRQL as ascertained by the EQ-5D visual analogue scale administered at four points in time was compared with the value predicted for each time point from a longitudinal model. The predictive model used available sociodemographic, physical, mental, and cognitive variables and the residuals were obtained by subtracting the observed from the predicted [43,44]. After centering the residuals to remove variability in magnitude, a group-based trajectory analysis was used to identify persons with common trajectories [45]. Seven patterns of residual fluctuations were identified. Sixty-seven percent of the 387 participants did not appear to undergo response shift while 15% showed a negative response shift, and 13% showed a positive response shift. Mayo et al have defined negative response shift over multiple points of time as an individual changing the scoring of his or her health from 'better than predicted' to 'worse than predicted'[43]. A hypothetical example of this is a person who, at 3 months after a stroke, has self rated health 20 points higher on a visual analogue scale (VAS) than predicted; at 6 months, the actual rating is only 5 points higher than predicted for that individual; and by 12 months, the VAS score is 10 points below predicted.

This approach was validated against the then test [43,44]. Further support for the approach came from the results of simulation studies in which no predominant trajectory emerged. This approach needs further work using predictors that are not assessed in the same cognitive frame as the outcome, e.g. measures of functional capacity such as tests of strength, walking speed and endurance, dexterity rather than self-report measures.

The advantage of this approach is that the response-shift pattern could be used to stratify a data set for post-hoc analyses to examine the effects of variables under study on change in the outcome among people who were identified as making no response shift. It would also be possible to identify predictors of a response shift group and even use the response shift group as an outcome if the intervention was such that a response shift was desired. This is a promising area of methodological development that requires further exploration.

Multivariate multi-level models

A common characteristic of studies of response shift in HRQL is data collection of multiple variables over time. An evaluation of 186 individuals with colorectal cancer at three time points using multivariate multi level models (an expansion of multiple regression) has identified the presence of reconceptualization [46]. With this method, a change in the regression parameters across times suggests reconceptualization. The outcome variable in this study was overall quality of life, while explanatory variables included the domains pain, energy, physical function, and mood. The relationships (regression parameters) between pain, energy and overall quality of life decreased significantly over time, while the relationship of physical function to overall quality of life increased significantly over time, suggesting reconceptualization response shift [46]. Further evaluation of this method and its comparison to other methods would be of significant value.

Rasch analysis

Rasch analysis has potential as a method for identifying possible response shifts when there is a change in the hierarchy of items in a measure over time. The hierarchy of items

relates to the difficulty of a particular item for the population assessed. In Rasch analysis and item response theory, differential item functioning (DIF) can be used to evaluate HRQL measures across different groups [47]. Differential item functioning evaluates the influence of a variable on the difficulty of an item [48]. If a measure contains items that demonstrate DIF, different groups respond to that item differently and measurement bias may occur [47]. If DIF is demonstrated across time in an item that is susceptible to response shift (such as perception of pain), the measurement bias due to DIF may be a result of a change in the meaning (response shift) of the item for individuals in the group.

Convergent validity

The various types of response shift are likely to be interdependent and may occur together, in parallel, or at the same rate [49]. Therefore, numerous studies have been designed to demonstrate convergence between methods. For example, the ability of the then test and SEM to detect response shift in cancer patients, using the SF-36 was compared and recalibration was noted using both approaches in bodily pain and role functioning due to physical problems [50]. Schwartz et al [30] used both the then test and longitudinal factor analysis in a five year follow-up of multiple sclerosis patients, which found response shift (recalibration and reconceptualization) in physical role limitations with both methods.

Ahmed et al [3,27,34] utilized SEM, the then test, and an individual measure (Patient Generated Index) in three studies with stroke survivors in the first six months post stroke. The then test demonstrated recalibration, the PGI showed reconceptualization and reprioritization, but SEM did not demonstrate response shift in that same time period [51]. It has been suggested that a reason for not finding response shift with the SEM

model could be related to the measurement model used [52]. Based on the results of the three methods, Ahmed provided recommendations for determining which method to consider when evaluating response shift in HRQL [51]. These recommendations are based on the memory and cognition of the participants, the number of participants available, and the type of response shift being evaluated [51]. Further studies which demonstrate whether convergent validity is present between the various methods of evaluating response shift are needed to strengthen this literature.

Clinical importance

A meta-analysis of response shift suggested that response shift was clinically important in some individual studies, but this was not apparent when studies were aggregated [53]. Mayo et al [43] demonstrated in work on different trajectories that the negative and positive response shift groups essentially cancelled themselves.

Many investigators have reported minimal important difference (MID) values to better interpret whether change scores for various HRQL instruments are clinically relevant [54]. The calculation of MID values is complicated if one takes response shift into account, as both the anchor-based and distribution-based approaches to measuring the MID assume response shift has not occurred [12]. As change in HRQL can be underestimated if response shift is not taken into account, the clinical importance of a change may also be underestimated. Accurate change scores are required to determine the MID.

Assuming that an MID is accurate, not accounting for response shift in a study may adversely affect the interpretation of HRQL change. For example, if the MID on a HRQL

measure is 7 points, a 5 point change (post test - pre test) would not be clinically relevant. However, if response shift is taken into account, using a then test approach, the difference between the then test and post test is 8 points (a 3 point response shift); the results of the intervention would now be considered clinically relevant. Response shift acts in determining the MID and whether a clinical relevant change in HRQL has occurred owing to an intervention or other variable under study.

Recommendations for Future Research Priorities

Researchers need to come to a generally accepted and clear definition of response shift, using prior work as a basis [13,40,55]. Researchers should also continue their attempt to empirically confirm the theoretical underpinnings of response shift by developing and testing theoretical models of response shift [13,23] such as Visser et al [22] have initiated. It is also encouraged that researchers analyze appraisal, including factors such as strategies used, standards of comparisons, and frames of reference used [23], including further evaluation of tools such as the Appraisal Profile which will be beneficial for understanding the measurement of health in individuals with disability [8].

As described in this paper, the terms ‘positive’ and ‘negative’ response shift have been defined differently by different authors. For example, Osborne et al and Mayo et al have described negative and positive response shift differently [9, 43]. It is suggested that a common usage of ‘positive’ and ‘negative’ response shift be adopted.

Further attention to the timing of response shift is needed to understand how long it takes for individuals in various health states to undergo a response shift. It is unknown how many data points a researcher should search for response shift [12]. Study of the timing

of response shift would also assist in the evaluation of theoretical models and help researchers and clinicians in determining appropriate timeframes for HRQL evaluation and interventions.

During a response shift workshop at the 14th International Society for Quality of Life Research conference in October 2007, numerous participants suggested that response shift research should link further with other methodologies and fields of research, such as qualitative methodologies and cognitive interviewing, adaptation psychology and cognitive psychology. Papers presented at this conference included initial work on new areas of response shift research including classification and regression tree models and differential item functioning to identify response shift [56,57].

There is another area of study to which response shift may be related; in 1999, Wilson suggested that response shift may be a type of placebo effect [16]. Both the placebo effect and response shift reflect changes that are not explained by natural history or recognized biological and physiological effects [16]. This deserves further evaluation.

Further evaluation of Current Methods and Investigation of Newer Methods

Instrument improvement of the older methods and development of promising new methods should be a focus of further research. It is unclear what type of response shift is evaluated with some of the methods currently used. This question also arises when evaluating new methods (such as residual analysis), and needs to be considered.

It is highly recommended that recall ability is analyzed when using the then test; if recall ability is inaccurate, the then test may not be an appropriate instrument to measure response shift. Ahmed determined recall ability by asking individuals a series of

questions related to their health care at baseline during the pretest and posttest evaluations [3]. Further research of other methods which can determine recall ability is encouraged.

We must also assess the reliability of the then test in regards to both test-retest and inter-rater reliability. It is also recommended that investigators come to a consensus on the wording of the then test and that it be reported in each research publication.

Development of new individualized methods which may be less burdensome than current methods will assist both individuals and researchers in the evaluation of response shift at this level. It will also assist in the understanding of the meaning of HRQL and response shift for individuals. Advances in the availability of computerized versions of individualized methods such as the SEIQoL [58] is also important to further explore.

There are numerous areas of study which could strengthen the use of SEM in response shift. Groups of individuals experiencing different health conditions and catalysts for change should be evaluated with the SEM method as described by Oort to allow further understanding of response shift over time in various health conditions [35]. The use of SEM with longitudinal data of greater than two time points will assist in determining the time-frame when response shift is most likely to occur in various conditions. This will be beneficial to describe the pattern of response shift and when it occurs, of which, little is known. Mayo recommends a minimum of 4 time points. It would be beneficial to determine the proportion of individuals in a group which need to demonstrate response shift so that factor changes are identifiable with SEM at the group level. This aspect needs further evaluation. Two methods are currently used to determine the presence of

response shift using SEM [33,35]. It would be beneficial to determine if similar patterns of response shift are suggested between the methods by assessing convergent validity in the same sample.

The use of residuals to identify those who potentially underwent response shift is an intriguing area of study. This would enable multi-group analysis in SEM with, for example, those who appear to have undergone response shift based on residuals, and those who do not appear to have undergone response shift. This is a promising area of study with secondary data analysis. The evaluation of DIF with Rasch analysis over time in conjunction with other methods to evaluate response shift will help to determine if this may be a useful method in assessing for response shift with secondary data analysis.

Further study of the convergent validity among various techniques, especially some of the newer and older techniques will help to strengthen our knowledge of how these techniques compare to each other. The interpretation of various studies may also be made easier.

Clinical importance

As previously described, the measurement of clinical importance has implications for interpreting the measurement of change both clinically and in research endeavors, and deserves future study. Techniques for measuring the MID could be incorporated into future studies assessing the recalibration response shift effect. Thus, studies could determine if calculating the MID after accounting for response shift changes interpretation of the MID.

On a slightly different vein, it is also important to study the views and perceptions of clinicians and researchers regarding whether they view response shift is an important concept to measure clinically. Views of clinicians and other researchers may inform response shift researchers about opportunities for knowledge translation of response shift research which is applicable in clinical and research settings.

Reporting guidelines

Authors of response shift need to “speak the same language” by having common reporting mechanisms. This will enable the readers to interpret and evaluate the response shift literature in a more straightforward manner. Reporting standards for response shift research have been suggested and we encourage that researchers embrace them [53]. A summary of the above mentioned research priorities is described in Table 2-1.

Conclusion

Response shift is a growing field of study in the area of quality of life and other patient-reported outcomes (PROs). There are many aspects which require further study and many new and exciting opportunities to explore. This paper has offered recommendations for future response shift research priorities : 1) Improve the theoretical underpinnings of response shift (e.g. come to a clear definition of response shift, empirically test theoretical models and develop new models of response shift if needed), 2) Continue to evaluate the validity of current and emerging methods, 3) Better understand how response shift and the MID are related and when response shift itself is clinically important, and 4) Utilize guidelines for reporting on response shift. It is hoped that this paper will also stimulate discussion amongst HRQL researchers and those interested in this field.

Acknowledgements

We wish to thank members of the Response Shift Special Interest Group of ISOQOL at the occasion of the 12th meeting for suggesting we undertake this review. We would also like to thank members of the Response Shift Special Interest Group of ISOQOL for helpful comments on early versions of this manuscript: Dr. Carolyn Schwartz, Dr. Sara Ahmed, Dr. Mechteld Visser, Dr. Lena Ring and Dr. Ida Korfage.

Reference List

1. FDA (2006). *Guidance for Industry: Patient Reported Outcomes: Use in Medical Product Development to Support Labelling Claims: Draft Guidance*. FDA, Rockville,MD.
2. Shimozuma, K., Imai, H., Ohsumi, S., & Ono, M. (2007). Recent Topics in Health Outcomes Research in Oncology. *Breast Cancer*, 14, 60-65.
3. Ahmed, S., Mayo, N., Wood-Dauphinee, S., Hanley, J., & Cohen, S. (2004). Response Shift Influenced Estimates of Change in Health-Related Quality of Life Poststroke. *Journal of clinical epidemiology*, 57, 561-570.
4. Razmjou, H., Yee, A., Ford, M., & Finkelstein, J. (2006). Response Shift in Outcome Assessment in Patients Undergoing Total Knee Arthroplasty. *The Journal of Bone and Joint Surgery*, 88-A, 2590-2595.
5. Ring, L., Hofer, S., Heuston, F., Harris, D., & O'Boyle, C. (2005). Response Shift Masks the Treatment Impact on Patients Reported Outcomes (PROs): the Example of Individual Quality of Life in Edentulous Patients. *Health and Quality of Life Outcomes*, 3, Retrieved May 21, 2007 from <http://www.hqlo.com/content/3/1/55>.
6. Mayo, N., Wood-Dauphinee, S., Ahmed, S., Gordon, C., Higgins, J., McEwen, S. et al. (1999). Disablement Following Stroke. *Disability and Rehabilitation*, 21, 258-268.
7. Anderson, C., Carter, K., Brownlee, W., Hackett, M., Broad, J., & Bonita, R. (2004). Very Long-Term Outcome After Stroke in Auckland, New Zealand. *Stroke*, 35, 1920-1924.

8. Schwartz, C., Andresen EM, Nosek, M., Krahn, G., & RRTC Expert Panel on Health Status Measurement (2007). Response Shift Theory: Important Implications for Measuring Quality of Life in People With Disability. *Archives of Physical Medicine and Rehabilitation*, 88, 529-536.
9. Osborne, R., Hawkins, M., & Sprangers, M. (2006). Change of Perspective: A Measureable and Desired Outcome of Chronic Disease Self-Management Intervention Programs That Violates the Premise of Preintervention / Postintervention Assessment. *Arthritis and Rheumatism*, 55, 458-465.
10. Schwartz, C. E. & Sprangers, M. A. (1999). Methodological Approaches for Assessing Response Shift in Longitudinal Health-Related Quality of Life Research. *Social Science and Medicine*, 48, 1531-1548.
11. Schwartz, C. & Sprangers, M. (2000). *Adaptation to Changing Health - Response Shift in Quality-of-Life Research*. Washington: American Psychological Association.
12. Schwartz, C., Sprangers, M., & Fayers, P. (2005). Response Shift: You Know it's There But How do You Capture it? Challenges for the Next Phase of Research. In P.Fayers & R. Hay (Eds.), *Assessing Quality of Life in Clinical Trials - Methods and Practice* (2 ed., pp. 275-290). Oxford University Press.
13. Sprangers, M. A. & Schwartz, C. E. (1999). Integrating Response Shift into Health-Related Quality of Life Research: A Theoretical Model. *Social Science and Medicine*, 48, 1507-1515.

14. Schwartz, C and Fairclough, D (2004, April). Emerging Areas of Quality of Life Research. *International Society for Quality of Life Research Newsletter*, 9, 4-5.
15. Norman, G. (2003). Hi! How are you? Response shift, implicit theories and differing epistemologies. *Quality of life research an international journal of quality of life aspects of treatment, care and rehabilitation*, 12, 239-249.
16. Wilson, I. B. (1999). Clinical Understanding and Clinical Implications of Response Shift. *Social Science and Medicine*, 48, 1577-1588.
17. Mayo, N., Wood-Dauphinee, S., Cote, R., Durcan, L., & Carlton, J. (2002). Activity, Participation, and Quality of Life 6 Months Poststroke. *Archives of Physical Medicine and Rehabilitation*, 83, 1035-1042.
18. Echteld, M., Deliens, L., Ooms, M., Ribbe, M., & van der Wal, G. (2005). Quality of Life Change and Response Shift in Patients Admitted to Palliative Care Units: a Pilot Study. *Palliative Medicine*, 19, 381-388.
19. Rees, J., Clarke, M., Waldron, D., O'Boyle, C., Ewings, P., & MacDonagh, R. (2005). The Measurement of Response Shift in Patients with Advanced Prostate Cancer and their Partners. *Health and Quality of Life Outcomes*, 3, Retrieved May 21, 2007 from <http://www.hqlo.com/content/3/1/21>.
20. Howard, G., Ralph, K., Gulanick, N., Maxwell, S., Nance DE, & Gerber, S. (1979). Internal Invalidity in Pretest-Posttest Self Report Evaluations and a Reevaluation of Retrospective Pretests. *Applied Psychological Measurement*, 3, 1-23.

21. Golembiewski, R., Billingsley, K., & Yeager, S. (1976). Measuring Change and Persistence in Human Affairs: Types of Change Generated by OD Designs. *Journal of Applied Behavioural Sciences*, *12*, 133-157.
22. Visser, M., Oort, F., van Lanschot, J., van der Velden, J., Kloek, J., Gouma, D. et al. (2005). Testing the Response Shift Model of Sprangers and Schwartz Empirically. *Quality of Life Research*, *14*, 2109.
23. Rapkin, B. D. & Schwartz, C. E. (2004). Toward a theoretical model of quality-of-life appraisal: Implications of findings from studies of response shift. *Health and quality of life outcome*, *2*, 14, Retrieved May 21, 2007 from <http://www.hqlo.com/content/2/1/14>.
24. Fayers, P., Langston, A., & Robertson, C. (2007). Implicit Self-Comparisons Against Others Could Bias Quality of Life Assessments. *Journal of clinical epidemiology*, *60*, 1034-1039.
25. Hagedoorn, M., Sneeuw, K. C., & Aaronson, N. K. (2002). Changes in physical functioning and quality of life in patients with cancer: response shift and relative evaluation of one's condition. *Journal of clinical epidemiology*, *55*, 176-183.
26. Dibb, B. & Yardley, L. (2006). How Does Social Comparison Within a Self-Help Group Influence Adjustment to Chronic Illness? A Longitudinal Study. *Social Science and Medicine*, *63*, 1602-1613.
27. Ahmed, S., Mayo, N., Corbiere, M., Wood-Dauphinee, S., Hanley, J., & Cohen, R. (2005). Using the Patient Generated Index to Evaluate Response Shift Post Stroke. *Quality of Life Research*, *14*, 2247-2257.

28. Jansen, S. J., Stiggelbout, A. M., Nooij, M. A., Noordijk, E. M., & Kievit, J. (2000). Response shift in quality of life measurement in early-stage breast cancer patients undergoing radiotherapy. *Quality of life research an international journal of quality of life aspects of treatment, care and rehabilitation*, 9, 603-615.
29. Sprangers, M., vanDam, F., Broersen, J., Lodder, L., Wever, L., Visser, M. et al. (2000). Response Shift and Fatigue: The Use of the Thentest Approach. In C.Schwartz & M. Sprangers (Eds.), *Adaptation to Changing Health - Response Shift in Quality of Life Research* (1 ed., pp. 137-151). Washington: American Psychological Association.
30. Schwartz, C. E., Sprangers, M. A. G., Carey, A., & Reed, G. (2004). Exploring response shift in longitudinal data. *Psychology and Health*, Vol 19, 51-69.
31. O'Boyle, C., McGee, H., & Browne, J. (2000). Measuring Response Shift Using the Schedule for Evaluation of Individual Quality of Life. In C.Schwartz & M. Sprangers (Eds.), *Adaptation to Changing Health - Response Shift in Quality of Life Research* (1 ed., pp. 123-136). Washington: American Psychological Association.
32. Korfage IJ, de Koning HJ, & Essink-Bot ML (2007). Response shift due to diagnosis and primary treatment of localized prostate cancer: a Then test and a vignette study. *Quality of Life Research*, 10, 1627-1634.
33. Schmitt, N. (1982). The Use of Analysis of Covariance Structures to Assess Beta and Gamma Change. *Multivariate Behavioural Research*, 17, 343-358.

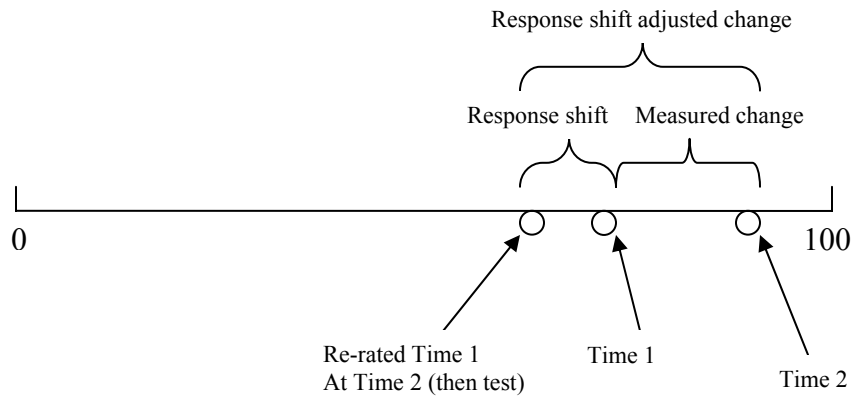
34. Ahmed, S., Mayo, N., Corbiere, M., Wood-Dauphinee, S., Hanley, J., & Cohen, R. (2005). Change in Quality of Life of People with Stroke Over Time: True Change or Response Shift? *Quality of Life Research, 14*, 611-627.
35. Oort, F. J. (2005). Using Structural Equation Modeling to Detect Response Shifts and True Change. *Quality of Life Research, 14*, 587-598.
36. Oort, F. J., Visser, M., & Sprangers, M. A. (2005). An Application of Structural Equation Modeling to Detect Response Shifts and True Change in Quality of Life Data from Cancer Patients Undergoing Invasive Surgery. *Quality of Life Research, 14*, 599-609.
37. Oort, F., van der Linden, Y., Sprangers, M., & Leer, J. (2005). Response Shift Detection in the Measurement of Symptom and Functional Status of Patients with Bone Metastasis. *Quality of Life Research, 14*, 2143.
38. Nolte, S., Elsworth, G., Sinclair, A., & Osborne, R. (2007). Replication of a Structural Equation Modeling Approach to Detect Response Shift in Participants of Health Education Programs. *Quality of Life Research* October[suppl.], A-13.
39. Barclay-Goddard, R., Lix, L., Mayo, N., Wood-Dauphinee, S., & Tate, R. (2007). A Framework for Assessing Response Shift in Post -Stroke Health Related Quality of Life Over Multiple Time Points. *Quality of Life Research* October[suppl.], A-14.
40. Oort, F. (2007). Response Shift and Measurement Bias. *Quality of Life Research* October[suppl.], A-14.

41. Schwartz, C. (1999). Teaching Coping Skills Enhances Quality of Life More Than Peer Support: Results of a Randomized Trial with Multiple Sclerosis Patients. *Health Psychology, 18*, 211-220.
42. Brossart, D., Clay, D., & Willson, V. (2002). Methodological and Statistical Considerations for Threats to Internal Validity in Pediatric Outcome Data: Response Shift in Self-Report Outcomes. *Journal of pediatric psychology, 27*, 97-107.
43. Mayo, N., Scott, S., Dendukuri, N., Ahmed, S., & Wood-Dauphinee, S. (2008). Identifying Response Shift Statistically at the Individual Level. In Press *Quality of Life Research*.
44. Mayo, N., Scott, S., Ahmed, S., & Dendukuri, N. (2005). Identifying Response Shift Statistically at the Individual Level. *Quality of Life Research, 14*, 1993.
45. Nagin, D. (1999). Analyzing Developmental Trajectories: A Semiparametric, Group-Based Approach. *Psychological Methods, 4*, 139-157.
46. Lowy, A. & Bernhard, J. (2004). Quantitative Assessment of Changes in Patients' Constructs of Quality of Life: An Application of Multilevel Models. *Quality of Life Research, 13*, 1185.
47. Lai, T., Teresi, J., & Gershon, R. (2005). Procedures for the Analysis of Differential Item Functioning (DIF) for Small Sample Sizes. *Evaluation and the Health Professions, 28*, 283-294.

48. Nijsten, T., Whalley, D., Gelfand, J., Margolis, D., McKenna, P., & Stern, R. (2005). The Psychometric Properties of the Psoriasis Disability Index in United States Patients. *Journal of Investigative Dermatology*, *125*, 665-672.
49. Schwartz, C. E. & Sprangers, M. (2000). Discussion: Implications of Response Shift for Clinical Research. In C.E.Schwartz & M. Sprangers (Eds.), *Adaptation to Changing Health - Response Shift in Quality-of-Life Research* (pp. 211-214). Washington: American Psychological Association.
50. Visser, M., Oort, F., & Sprangers, M. (2005). Methods to Detect Response Shift in Quality of Life Data:A Convergent Validity Study. *Quality of Life Research*, *14*, 629-639.
51. Ahmed, S., Mayo, N., Wood-Dauphinee, S., Hanley, J., & Cohen, S. (2005). The Structural Equation Modeling Technique Did Not Show a Response Shift, Contrary to the Results of the Then Test and the Individualized Approaches. *Journal of clinical epidemiology*, *58*, 1125-1133.
52. Borsboom, G., Korfage IJ, Essink-Bot ML, & Duivendoorn, H. (2007). Letter to the Editor - The Structural Equation Modeling Technique Did Not Show a Response Shift, Contrary to the Results of the Then Test and the Individualized Approaches. *Journal of clinical epidemiology*, *60*, 426-428.
53. Schwartz, C., Bode, R., Repucci, N., Becker, J., & Sprangers, M. (2006). The Clinical Significance of Adaptation to Changing Health: A Meta-Analysis of Response Shift. *Quality of Life Research*, *15*, 1533-1550.

54. Osoba, D (2006). ISOQOL's 2006 President Award's Address. *International Society for Quality of Life Research Newsletter*, 11, 6-9.
55. Oort, F., Visser, M., Hartman, E., Locadia, M., Niewkerk, P., & Sprangers, M. (2005). Formal Definitions of Response Shift in Measurement and Conceptual Perspectives. *Quality of Life Research*, 14, 2109.
56. Rapkin, B., Li, Y., & Patel, S. (2007). Classification and Regression Tree Analysis to Support QOL Theory Development. *Quality of Life Research* October[suppl.], A-35.
57. Schwartz, C., Martha, J., Swaim, B., Bode, R., & Kim, D. (2007). Detecting Response Shift in Using Rasch Analysis of Then Test Data. *Quality of Life Research* October[suppl.], A-15.
58. Ring, L., Lindblad, K., Bendtsen, P., Vicklund, E., Jansson, R., & Glimelius, B. (2006). Feasibility and validity of a computer administered version of SEIQoL-DW. *Quality of Life Research*, 15, 1173-1177.

Figure 2-1 Then test



This figure represents a measure of HRQL on a 0-100 scale with a traditional pre-test at Time 1, post-test at Time 2, and a then test occurring also at Time 2.

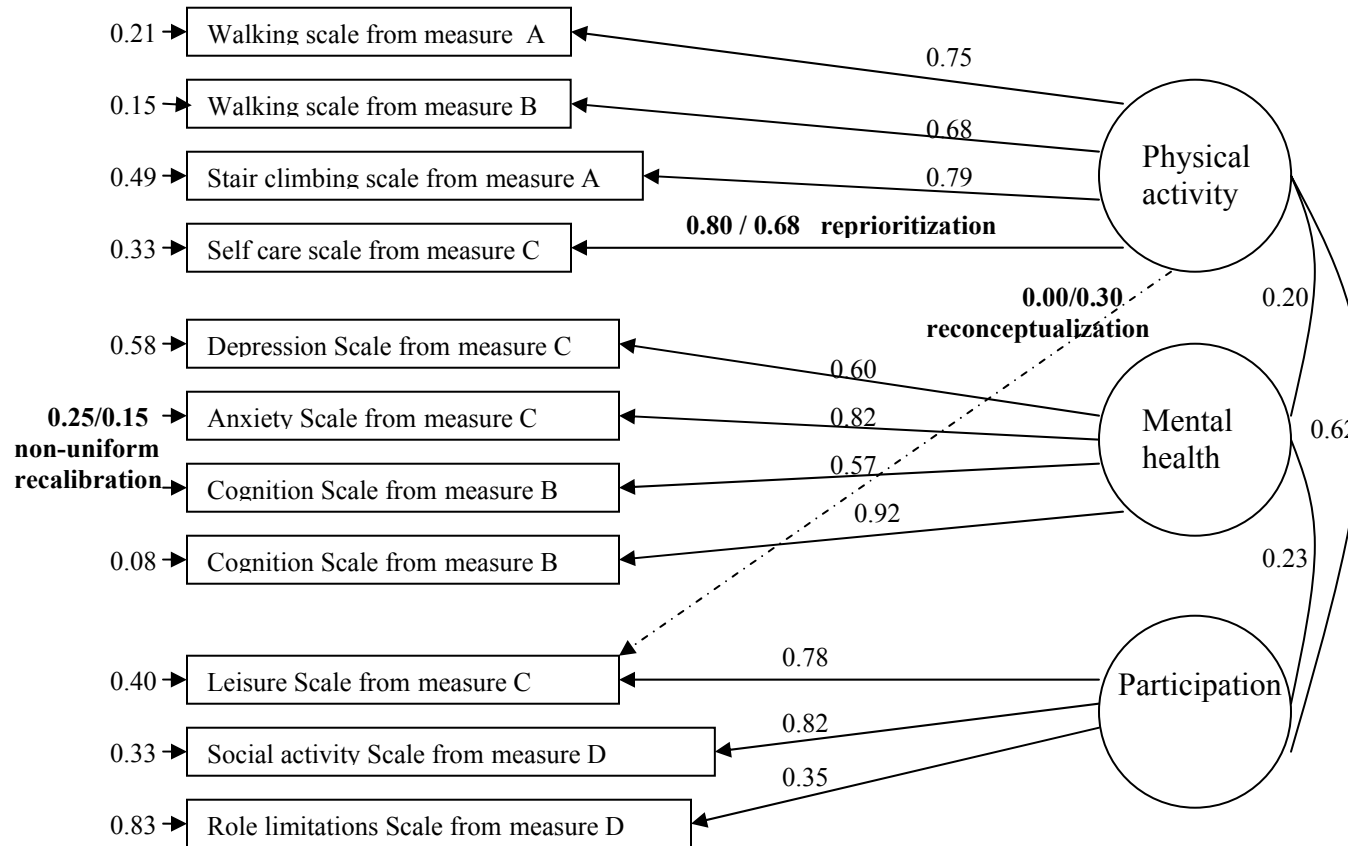
Example of a typical then test question: Please remember to when you last filled out this questionnaire 1 month ago. Please provide a new judgment about your quality of life at the time you last filled out this questionnaire 1 month ago. Rather than try to recall your prior responses, please offer a renewed judgment of your quality of life then, given what your perspective is today.

Figure 2-2 The Patient Generated Index (PGI)

STEP 1	STEP 2	STEP 3
Select 5 areas	Rate areas on ability (0-10 points)	Spend 12 points reflecting importance of each area(weight)

Final score (/100) = the extent to which reality falls short of patients' hopes and expectations for those areas of life for which they would most value an improvement.

Figure 2-3 Factor Analysis: A Hypothetical Measurement Model of HRQL



Hypothetical measurement model with 3 latent variables (Physical Activity, Mental Health, Participation) associated with measures variables from 4 hypothetical outcome measures (A,B,C,D). Changes over time in error variances suggests non-uniform recalibration, changes in factor loading values reflects reprioritization, changes in factor loading patterns reflects reconceptualization, and uniform recalibration is reflected in changes in the intercepts (not demonstrated in this picture). Based on Oort [35].

Table 2- 1 Summary of Recommendations of Future Research Priorities

Definition and Theoretical Models of Response Shift	Further Evaluation of Current Methods and Investigation of Newer Methods	Clinical Importance	Reporting Guidelines
Agree on clear definition of response shift	Analyze recall ability of the then test	Evaluate clinical importance of response shift	Follow suggested reporting standards [53]
Empirically test theoretical models of response shift	Assess the reliability of the then test	Incorporate MID into future studies of recalibration response shift	
Evaluate appraisal, strategies used, standards of comparisons, and frames of reference	Come to a consensus on the wording of the then test	Study the views and perceptions of clinicians and researchers re: clinical importance of response shift	
Adopt common terminology for ‘positive’ and ‘negative’ response shift	Development of less burdensome individualized methods		
Determine timing of response shift in various health states	Evaluation of response shift over at least 4 time points with SEM		
Link response shift research with other fields of research	Convergent validity between various methods		
	Use of residual analysis to inform multi-group SEM analysis		
	Evaluation of DIF over time with Rasch analysis to determine if useful in identifying response shift		

Chapter 3: Study Methodology and Descriptive Data Analysis

Summary of chapter 3

This chapter discusses the rationale of the current study and provides objectives and hypotheses for the entire thesis. The data source and analysis of data are described.

Detailed description of the characteristics of the study participants and variables utilized follows. The information here is described in more detail than is possible or necessary to be included in manuscripts 2 and 3.

Rationale for the thesis

Quality of life and health related quality of life are clearly important to consider for understanding recovery following stroke. Chapter 1 provided compelling data that stroke is a condition with major functional sequelae and that both the process of recovery and rehabilitation can induce a response shift in global outcomes such as HRQL. Ahmed et al. are the only group who has studied response shift in stroke. Using an individualized approach to identifying important QOL constructs over time, and using the then test, response shift was identified^{1,2}. However, using an SEM approach with the SF-36 as the measurement model, response shift was not identified³.

The discrepancy in identifying response shift using different methods may arise from the constructs studied, the statistical approach applied, or the timing of the reassessments with the use of SEM. The method used by Ahmed et al., based on the work of by Schmitt, involved confirmatory factor analysis (CFA) to identify the presence of recalibration and reconceptualization response shift^{3,4}. Oort (2005) has demonstrated an alternative CFA method specifically to determine all types of response shift in HRQL⁵. The strength of

this method is that it can be integrated with the response shift model of Sprangers and Schwartz (1999) and it describes all 3 types response shift ⁶. The method described by Oort has been evaluated with individuals with cancer who have had extensive surgery for cancer and individuals undergoing a self management course for arthritis ^{7, 8}. It has been suggested that as SEM analysis is based on estimates at a group level, some individuals may have shown response shift, but this would have been hidden by the group level analysis which occurs with SEM ^{3, 5}. Interestingly, Mayo et al, using this same data set and another statistical approach found that there were subgroups of the population who did show a response shift (both positive and negative) but the majority of subjects (67%) showed no response shift ⁹.

Ahmed et al, evaluated response shift between six weeks and six months post stroke ³ but a longer time-frame may be needed to detect response shift in different constructs.

Schwartz et. al. (2004) suggested that recalibration was more likely in the first few months after a threatening event and that clients with more severe symptoms engage in recalibration for longer time intervals than those with milder symptoms – perhaps response shift will be more evident between one and 12 months, rather than between one and six months.

Ahmed also examined response shift using only the SF-36 ³. It has been suggested that the less than optimal fit of the SF-36 based measurement model in the original study may have affected the evaluation of response shift; a better fitting model may have had a different outcome ¹⁰. In addition, the SF-36 is a composite of several constructs, which themselves may be subjected to response shift in differing direction and times post-stroke. It has been suggested that, “Future work, particularly in stroke, needs to go

beyond the SF-36 and include measures of other important constructs related to quality of life...”¹¹. This study intends to do just that: go beyond the SF-36 and use separate models of the constructs physical function, mental health, and participation.

Global Aim

The global aim of this study is to contribute to the development of a model of the dynamics and trajectory of change in health related quality of life (HRQL) post-stroke.

Overall Objective

The overall objective is to assess response shift in construct specific HRQL models.

Specific Objectives

1. To estimate if there is a difference in the factor structure of the latent constructs (physical health, mental health, participation) of HRQL at different points in time post stroke.
2. Estimate whether time contributes to the factor structure of the latent constructs (physical health, mental health, participation) of HRQL (using a latent covariance model to determine reconceptualization, reprioritization, and non uniform recalibration).
3. Estimate whether time contributes to the mean pattern of the latent constructs (physical health, mental health, participation) of HRQL (using a latent mean model to determine uniform recalibration).

Hypotheses

1. The factor structure of the latent construct mental health is not invariant over time. The mean pattern of the latent domain mental health is not invariant over time. Response shift will be detected.
2. The factor structure of the latent construct physical function is invariant over time. The mean pattern of the latent domain physical function is invariant over time. Response shift will not be detected.
3. The factor structure of the latent construct participation is not invariant over time. The mean pattern of the latent domain participation is not invariant over time. Response shift will be detected.
4. Response shift will be seen at 12 months post stroke.

Source of data

The data for this examination of response shift came from a longitudinal study of health related quality of life post stroke - “Understanding Quality of Life Post-Stroke: A Study of Individuals and their Caregivers”¹². This flagship project funded by the Canadian Stroke Network, has accumulated data from 678 persons with stroke and 410 informal caregivers in Montreal, Toronto and London at 1, 3, 6, and 12 months post stroke. Three hundred and forty-two subjects have complete data at all 4 time points. The main investigators were Sharon Wood-Dauphinee (Principal Investigator) and Nancy Mayo, Mark Bayley, Angela Cheung, Jayne Garland, and Jeffrey Jutai.

The data arising from this study was ideally suited for studying response shift because it included variables covering all of the major functions and disabilities related to stroke, as defined by the International Classification of Functioning (ICF). There is also data on personal and environmental factors. See Table 3-1 for a description of measures fitting these categories.

Table 3-1 Data Source Variables

Impairments (Measure)	Activity limitations (Measure)	Participation restrictions (Measure)
Stroke severity (Canadian Neurological Scale, CNS) ¹³ type of stroke, co-morbidities (Charlson Index ¹⁴), mental status (Telephone Mini Mental Status Exam, MMSE) ¹⁵ Selected subscales of HRQL indices	Activities of Daily Living (Barthel Index) ¹⁶ Selected subscales of HRQL indices	Selected subscales of HRQL indices
Personal factors: gender, age, language, country of origin, education, occupation		
Environmental factors: Psychosocial Impact of Assistive Devices Scale ¹⁷ , with whom the stroke survivor lives, number of children, medication, length of hospital stay, OARS - Social Resources ¹⁸		
HRQL generic: Short-Form 36 (SF-36) ¹⁹ , EuroQuol (EQ5D) ²⁰ , Health Utilities Index (HUI) ²¹ .		
HRQL stroke specific: Stroke Impact Scale (SIS) ²² , Preference-Based Stroke Index (PBSI) ²³		

Supplemental Table 3-1 provides additional information on each generic and stroke specific HRQL index used, including content, items, response options, and psychometric properties.

Figure 3-1 demonstrates how the HRQL indices, constructs of HRQL, and ICF categories are intertwined and connected. The generic and stroke specific indices of HRQL provide

items and /or subscales that reflect each of the constructs of HRQL (mental health, physical function and participation). Each of the constructs relate to one or more aspects of disability as defined by the ICF (impairments, activity limitations, or participation restrictions).

Figure 3-1 Association of the ICF, HRQL constructs and HRQL indices

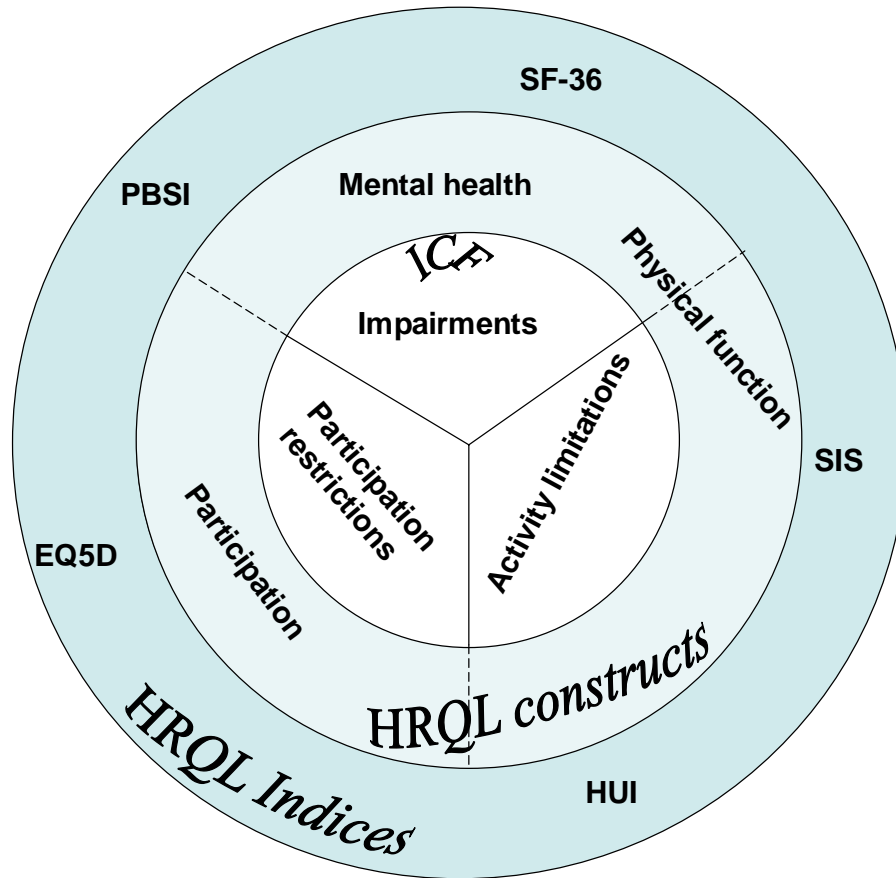


Table 3-2 lists the items and subscales which describe the HRQL constructs of physical function, mental health, and participation as well as overall health and wellbeing. This formed the basis of developing the measurement models for each construct.

Table 3-2 Items and subscales from HRQL measures and related constructs

Measure	Physical function	Mental health	Participation	Overall health
SF-36	Physical function Body Pain	Vitality Mental health	Social function Role physical Role emotional	General health
EQ5D	Mobility Self care Pain	Anxiety	Usual activities	VAS - thermometer
HUI	Self care Ambulation Dexterity Pain	Emotion Cognition		
SIS	Hand function Mobility ADL&IADL	Emotion Memory	Social participation	VAS- Global recovery
PBSI	Physical activity / sports Stairs Walking	Self esteem Coping Memory	Driving Work Recreational activities	

SF-36 = Short Form 36, EQ5D = EuroQol, HUI = Health Utilities Index, SIS = Stroke Impact Scale, PBSI = Preference Based Stroke Index, VAS = visual analogue scale, ADL = activities of daily living, IADL = instrumental activities of daily living

Analytical Approach

Table 3-3 summarizes the analytical approach to both characterizing the study participants, characterizing variables, determining patterns of missing data, analyzing missing data, developing measurement models for each construct, and identifying response-shift susceptible constructs.

Table 3-3 The Analytical Approach

Task	Method	Software
Characterizing study participants	Descriptive stats, cross-sectional at 4 points in time, missing data excluded: personal factors, disability related factors, HRQL indices	SAS 9.14 ²⁴
Characterizing variables	Univariate tests for normality of the items or subscales for HRQL indices	SAS 9.14
Missing data evaluation	χ^2 comparison of specific outcomes and missingness to determine missing data pattern	SAS 9.14
Parameter estimation of models	Full information maximum likelihood (FIML)	LISREL 8.72 (SIMPLIS & PRELIS) ²⁵
	Multiple imputation	SAS 9.14
	Robust maximum likelihood estimation	LISREL 8.72
Mental health model building	Based on Medical Outcomes Study Framework of Health Indicators ²⁶ and incorporating modification indices	LISREL 8.72
Physical function model building		
Participation model building		
Identification of response shift	SEM approach described by Oort (2005) to identify reconceptualization, recalibration, or reprioritization response shift	LISREL 8.72

Analyses for missing data

Data was analyzed to determine the type of missingness. Missing data is an issue with SEM. The type of missing data needs to be determined, and appropriate methods used to estimate missing values based on the reason for the missing data^{27, 28}. Missing data is common in longitudinal studies for a variety of reasons: decreased observations may lead to decreased power, HRQL will be overestimated if data that are missing are from clients

with worse outcomes, and those with a different number of missed assessments may have different HRQL²⁹.

We compared missingness in the mental health and physical health summary of the SF-36 and the EQ5D VAS with age (70+, <70), gender, initial stroke severity (Canadian Neurological Scale 6+, <6)³⁰, and Barthel score at admission (100, <100)^{16, 31}. A χ^2 was completed for each comparison at each time point.

Full information maximum likelihood (FIML) was the estimation method used for SEM analysis, to account for missing data³². With FIML, missing data is not imputed; parameter estimates are estimated directly from the raw data with iterative computer algorithms³². The parameter estimates that are produced with FIML are considered to be unbiased under the assumption of missing at random; FIML also assumes normality³². SEM assumes missing at random or missing completely at random data pattern³².

Characterizing variables (non-normality)

Using the robust maximum likelihood estimation (MLE) method helps to account for non-normality in data³³. The Satorra-Bentler χ^2 is the commonly used fit statistic with the robust method; it is not available in conjunction with FIML in LISREL 8.72. When data is continuous or ordered categorical non-normal data, the Satorra-Bentler χ^2 performs better than the χ^2 from other, non-robust methods³³. All measured variables used in the mental health, physical function, and participation models were analyzed for normality. If data is moderately non-normal, a Satorra-Bentler scaling method is suggested³³. The papers in chapters five and seven, “A Framework for Testing Response Shift in Health Related Quality of Life Over Multiple Time Points – A Demonstration

with Individuals Post Stroke” and “Can Response Shift Occur in Physical Function? Using Stroke as a Model” describe sensitivity analyses comparing FIML and robust MLE to determine if either method is appropriate to use given both missing data and non-normality of data.

Latent curve models

Latent curve models were developed with selected outcomes to determine if observed change across the four time points was linear. A model was developed with each selected outcome at each of the four time points as the measured variables; the intercept and slope were modeled as the latent variables. Each model was determined with free error variances and also with equal error variances. Equal error variances are usually assumed with longitudinal data, as influences are assumed to be the same and the same indices are used across time³⁴. A well fitting model suggests linear change over time. The selected outcomes included: the mental health and physical health components of the SF-36, the EQ5D VAS feeling thermometer, summary scales of HUI3 multiattribute score, PBSI summary score, SIS physical summary score, and the EQ5D multiattribute score

Building Models of the HRQL constructs

The models of HRQL used for this study are models of the following constructs: physical function, mental health, and participation. Choices of constructs reflect the common inclusion of physical health, mental health and social functioning in HRQL measures³⁵. Social functioning has been redefined as the broader construct of participation, based on the ICF definition³⁶. Content of the constructs is based in part on the Medical Outcomes Study Framework of Health Indicators²⁶. It has been stated that “Realizing the full

benefit of SEM requires integrating advanced statistical techniques with carefully considered formalizations of substantive knowledge and theory³⁷. It is therefore important that the models used are based on theory and modified only when sensible based on both theory and experience.

Physical health (also referred to as physical function) consists of: self care / ADL , moving and walking (including U/E and L/E function and transfers) , pain , energy / fatigue , and sleep ²⁶. Mental health consists of: cognitive functions (e.g. forgetfulness, difficulty concentrating) and psychological distress / wellbeing (e.g. anxiety, depression, emotional control, and positive affect, feelings of belonging) ²⁶.

Participation is defined as ‘involvement in a life situation’ ³⁶. Participation includes aspects such as: social function, role function, role limitations due to emotional problems or to physical health ²⁶. Examples of participation from HRQL measures in the study include: usual activities, recreation activities, work, and driving.

The initial model building process for all models included: model conceptualization, parameter identification, parameter estimation, data-model fit assessment, and model modification ³⁸. Goodness of fit assessment was described in Chapter 1.

Determining response shift

The initial cross sectional models were analyzed with pairwise comparisons at each time period to informally evaluate comparisons of factor loadings. In the longitudinal models, all four time points are represented, with residuals correlated between times ⁵. Steps were followed as described by Oort to identify localization of response shift for hypotheses 1-3 ⁵. Step 1 includes the establishment of an appropriate measurement model. There are no

constraints in this model. Step 2 involves fitting a model of no response shift. In this model intercepts, factor loadings and residual factor variances are all constrained across times. In this step, a significant chi-square difference between the two models suggested response shift somewhere in the model. In step 3, the identification of a specific type and timing of response shift is undertaken. Constraints are removed step by step to determine the type(s) of response shift present. Step 4 is the final model. In this model, constraints are added to determine other types of change. Each step is evaluated with a chi-squared difference test. The main focus of this research was in determining response shift type and location, up to and including step 3.

Written approval to use the dataset has been received from the principal investigator, Dr. Sharon Wood-Dauphinee (McGill University) and Ethics approval was granted from the Research Ethics Board at the University of Manitoba.

Results

Characteristics of persons

Descriptive statistics describing the participants are presented in Supplemental Table 3-2. Supplemental Table 3-3 provides mean, median, and standard deviation for all HRQL measures used in the study.

Pattern of missing data

As seen in Supplemental Table 3-4, there was loss to follow up and missing data at each time point. Of the 678 participants in the study, 344 (50.4%) have data at all 4 time points. Sixty-seven percent of participants completed the study. Missingness was not

associated with age or gender, and was associated with more severe stroke severity and lower function (Barthel) only at time 1. Missingness was therefore determined to be missing at random (conditionally missing). See Table3-4. In discussion with Dr. Ken Bollen at the Summer Programme in Data Analysis course in 2006, he felt that since severity is controlled for over time, the data likely fits the assumption of missing at random (conditionally missing).

Table 3-4 Factors Associated with Missingness

Variable	MCS & PCS	EQ5D - VAS
Age	no	no
Gender	no	no
Stroke severity (Canadian Neurological Scale)	CNS<6 (more severe) Time1 only	CNS<6 (more severe) Time1 only
Function (Barthel)	Barthel<100 (lower function) Time1 only	no

MCS = mental health summary SF-36, PCS = physical health summary SF-36

Characteristics of the variables

All measured variables used in the mental health, physical function, and participation models were determined to be moderately non-normal with a skewness of <2 and kurtosis of <7³³. See Supplemental Table 3-5. All assessments of normality were completed at time one (1 month).

With moderately non-normal data, a Satorra-Bentler scaling method is suggested³³. The Satorra-Bentler statistic adjusts the χ^2 downward by an amount reflecting the kurtosis that is observed³⁹. Moderately non-normal data that is continuous or ordered categorical (with as little as three categories) can be analyzed with Satorra-Bentler scaling³³. The papers in Chapters five and seven, “Response Shift was Identified in Health Related Quality of Life Using a Framework Developed for Multiple Time Points: A Demonstration with Individuals Post Stroke” and “Can Response Shift Occur in Self-Perceived Physical Function? Using Stroke as a Model” describe sensitivity analyses comparing FIML for missing data and robust MLE for non-normal data.

Latent curve models

The mental health and physical health components of the SF-36 and the EQ5D VAS were found to show linear change over all time points. The latent curve models of linear change were well fitting for those summary scales; varying intercepts and rates of change were also noted. See Table 3-5. The models for the SF-36 had better fits with the fixed error variances, not statistically significantly different than the free models. The EQ5D visual analogue scale (VAS) model with free error variances was statistically significantly better than the fixed model, suggesting differing influences across time. Summary scales of HUI3 multiattribute score, PBSI summary score, SIS physical summary score, and the EQ5D multiattribute score did not have well fitting models for linear change, suggesting that observed change was in a pattern other than linear.

Table 3-5 Latent Curve Models Suggesting Linear Change

Model	Error variances	χ^2	df	Normed χ^2	p-value	RMSEA
SF-36 mental health summary	fixed	18.3	8	2.3	0.02	0.057
SF-36 physical health summary	fixed	20.3	8	2.5	0.01	0.062
EQ5D VAS	free	11.4	5	2.3	0.05	0.056

df = degrees of freedom, RMSEA = Root mean square error of approximation

Supplemental Tables

Supplemental Table 3-1 HRQL Measures

Measure	Stroke Impact Scale (SIS)	Short Form 36 (SF-36)	EuroQuol (EQ-5D)	Health Utilities Index (HUI)	Preference Based Stroke Index (PBSI)
Type	Disease specific to stroke	Generic health profile	Generic Utility and single index health status	Generic Utility	Stroke specific preference based
Purpose	Evaluative	Evaluative	Discriminative – provides a descriptive profile	Discriminative	Discriminative – stroke specific health index
Domains	8 Strength, hand function, mobility, ADL and IADL, emotion, memory, communication, social participation (Strength, hand function, mobility, ADL and IADL can be combined as a physical function domain)	8 physical functioning, role limitations due to physical, bodily pain, vitality, social functioning, role limitations due to emotional, mental health, general health	5 mobility, self care, usual activities, pain / discomfort, anxiety / depression visual analogue scale (VAS) assesses self perceived health status	9 (HUI 2/3) emotion, cognition (memory, thinking), self-care, pain / discomfort, vision, hearing, speech, ambulation, dexterity ²¹	10 Walking, stairs, physical activities, recreational activities, work / activity, driving, memory, speech, coping, self-esteem ²³
Number of items	Global perception of recovery question 59 (SIS 3.0) 64 (SIS 2.0) ²²	36	5	41	10
How administered	Interview – in person Mail version ⁴⁰	Interview - in person or phone Self completed	Interview –phone Self completed	Interview - in person or phone Self completed	Interview or mailed
Time	15-20 minutes	5-10 minutes	2-3 minutes	8-10 minutes, 20 for stroke patients	‘Short’
Scaling	5 point scale	2 points to 6 points depending on item	3 point scale	4 to 6 points depending on item	3 point scale

Measure	Stroke Impact Scale (SIS)	Short Form 36 (SF-36)	EuroQuol (EQ-5D)	Health Utilities Index (HUI)	Preference Based Stroke Index (PBSI)
Scoring	<p>Domains 0-100</p> <p>Global perception of recovery question 0-100</p> <p>higher = better</p> <p>algorithm required</p>	<p>Domains 0-100</p> <p>higher = better</p> <p>Physical and mental health components 0-100 with mean = 50, SD=10</p> <p>algorithm required</p>	<p>Each combination of scores describes a health state which is represented as a utility score</p> <p>0 = worst possible health state</p> <p>1 = best possible health state</p> <p>VAS = 0-100, higher = better</p> <p>for each item, lower score = better</p> <p>algorithm required</p> <p>245 possible health states⁴¹</p>	<p>Each combination of scores describes a health state which is represented by the global multi attribute score (utility score)</p> <p>0 = worst possible health state</p> <p>1 = best possible health state</p> <p>a score of less than 0 = state worse than death^{21, 42, 42}</p> <p>single summary scores for each attribute can also be calculated²¹</p> <p>algorithm required</p> <p>972,000 possible health states in HUI 3⁴²</p>	<p>A preference –weighted cumulative score is derived.</p> <p>0 = worst stroke scenario</p> <p>1 = best stroke scenario</p> <p>for each item, lower score = better</p> <p>59,049 possible health states</p> <p>A multi-attribute utility function is under development.</p> <p>(Poissant et al., 2003)</p>

Measure	Stroke Impact Scale (SIS)	Short Form 36 (SF-36)	EuroQuol (EQ-5D)	Health Utilities Index (HUI)	Preference Based Stroke Index (PBSI)
Proxy	Patient and proxy scores different, but bias was determined to be not clinically significant. Use caution in using proxy for subjective domains and be aware of potential bias. ⁴³	Elderly had poor to moderate agreement on all SF-36 subscales in 2 different settings and 2 types of proxies (health care professional and family/friend) ⁴⁴ Poor proxy agreement in stroke ⁴³ referring to Segal and Schall 1994	Likely useful for mobility and self care domains – use caution with proxy assessment in social function, pain, and overall HRQL. Depression / anxiety least reliable as proxy. ⁴⁵	It is suggested that caregivers complete the assessment for individuals with stroke if they are unable to do so. Moderate to high correlations and ICCs were found for each domain in the HUI2 and HUI3, even in subjective domains. ²¹	Stroke survivors and caregivers assisted in the development of preference weights, which did not differ significantly between groups. (Poissant et al., 2003)
Validity	Cross sectional construct validity – convergent and known groups has been demonstrated. Longitudinal construct known groups – adequate sensitivity in minor and moderate strokes at 1-3 and 1-6 months.	Cross sectional construct validity – convergent showed physical function subscale correlated higher with EQ5D physical domains than with other domains in stroke. Longitudinal construct not noted for stroke	Cross sectional construct validity – convergent showed high correlation with other measures of same domains in numerous diagnoses, including stroke Longitudinal construct not noted for stroke	Cross sectional construct validity - stroke patients had worse utility scores than those with arthritis ⁴² Longitudinal construct not noted for stroke	Cross sectional construct validity (convergent validity) showed high correlation between PBSI and most SF-36 subscales. Moderate correlation between the PBSI and the EQ5D. Known groups validity showed lower PBSI scores for those with severe stroke and higher PBSI scores for those with very mild stroke.(Poissant et al., 2003)
Reliability	Test-retest reliability generally good except for emotion ²²	Test-retest reliability generally good except for mental health ⁴⁶	Test-retest reliability generally good except for mental health ⁴⁶	Test-retest of general population was fairly good at one month for utility score. On domains of HUI 2, lowest reliability seen on emotion and pain	Test-retest reliability is to be evaluated in the future

Measure	Stroke Impact Scale (SIS)	Short Form 36 (SF-36)	EuroQuol (EQ-5D)	Health Utilities Index (HUI)	Preference Based Stroke Index (PBSI)
Responsiveness and Clinically meaningful change	In SIS 2.0, it is suggested that 10-15 points is clinically meaningful for most domains	Use of the SF-36 and EQ5D to assess change in individual patients over time was not supported by Dorman et al, 1998	Use of the SF-36 and EQ5D to assess change in individual patients over time was not supported by Dorman et al, 1998	Change of 0.03 is considered clinically important on the global utility scale ⁴² Not noted specifically for stroke	Not noted
Used in response shift studies?	Not noted	Yes _{3, 7}	Yes ₁	Not noted	Not noted
Other versions	SIS-16 (4 domains of physical function only)	SF-12 – still 8 domains, 12 questions SF-8 - 8 questions	Not noted	HUI 1, HUI 2, HUI3	Not noted

Note -if not otherwise stated, reference is ⁴⁷

Supplemental Table 3-2 Description of the Cohort

	n	Frequency	Percent
Female	677	305	45.1
Age at Stroke	676		
<45		63	9.3
45-54		73	10.8
55-64		131	19.4
65-74		170	25.2
75-84		175	25.9
>=85		64	9.5
<i>Mean and SD</i>	676	67.3	14.8
First Stroke	657	616	93.7
Side of Lesion	654		
Left		293	44.7
Right		325	49.5
Bilateral		36	5.5
Location of Hemiplegia	464		
Lower Extremity		24	5.2
Upper Extremity		56	12.1
Upper and Lower Extremities		384	82.8
Canadian Neurological Scale	667		
severe: 0-5		94	14.1
moderate-high: 5.5-9		260	39.0
moderate-low: 9.5-10.5		174	26.1
mild: 11-11.5		139	20.8
<i>Mean and SD</i>	667	8.5	2.5

Supplemental Table 3-3 Outcomes at all Times

Variable	1 month			3 months			6 months			12 months		
	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD
CNS	667	8.5	2.5									
Mini-mental score	471	19.5	3.3	403	20.1	2.6	381	20.6	2.2	384	20.6	2.2
Mini-mental/100	502	88.9	14.8	434	91.3	12.0	399	93.1	10.3	394	93.7	10.1
Barthel Index	668	74.5	25.9									
Stroke Impact Scale (SIS)												
Memory	609	83.7	21.7	530	86.4	19.2	477	87.5	18.1	444	88.1	18.3
Emotion	607	77.9	18.0	522	79.8	17.1	473	81.4	16.2	439	81.7	16.5
Communication	613	86.7	20.9	529	90.1	19.0	478	90.7	16.9	444	91.0	17.7
Strength*	606	64.8	29.0	527	70.7	26.1	478	71.6	25.9	442	73.9	25.4
Hand *	603	59.6	39.1	527	70.0	34.7	476	74.3	32.6	443	76.1	31.9
Mobility *	612	64.0	31.3	530	75.2	25.0	478	77.8	23.7	444	79.9	22.2
ADL*	611	66.7	27.1	529	77.1	22.9	480	78.9	22.0	445	81.3	21.5
Participation	610	52.0	29.3	528	64.9	27.9	477	69.9	27.1	445	75.1	25.9
SIS Physical*	612	64.3	28.0	530	74.3	23.3	480	76.7	22.4	445	78.9	21.4
SIS Recovery	598	60.3	27.3	511	70.0	23.3	470	72.0	22.4	438	76.0	21.9
Health Utilities Index version 3 (HUI3)												
Vision	509	0.92	0.16	464	0.93	0.13	439	0.94	0.12	424	0.92	0.15
Hearing	512	0.98	0.11	485	0.97	0.11	444	0.97	0.10	416	0.97	0.10
Speech	546	0.94	0.17	504	0.97	0.13	456	0.97	0.13	437	0.97	0.11
Ambulation	529	0.63	0.40	492	0.76	0.32	460	0.79	0.29	436	0.80	0.30
Dexterity	548	0.75	0.34	502	0.85	0.27	465	0.84	0.28	438	0.85	0.28
Emotion	520	0.86	0.24	481	0.91	0.17	442	0.92	0.16	424	0.91	0.19
Cognition	535	0.87	0.22	502	0.89	0.20	462	0.89	0.20	435	0.88	0.20
Pain	537	0.84	0.29	500	0.85	0.27	464	0.85	0.27	435	0.84	0.29

Supplemental Table 3-3 Outcomes at all Times (continued)

Variable	1 month			3 months			6 months			12 months		
	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD
HUI3 weighted	418	48.9	37.6	401	62.0	32.5	379	63.6	31.8	375	62.6	33.5
Health Utilities Index Version2 (HUI2)												
Self care	504	0.64	0.47	469	0.77	0.41	439	0.80	0.39	436	0.82	0.37
Euroquol (EQ-5D)												
Mobility	530	1.7	0.7	463	1.5	0.5	420	1.5	0.6	413	1.2	0.5
Self care	529	1.5	0.6	461	1.3	0.5	420	1.3	0.5	413	1.2	0.5
Usual activities	529	2.0	0.7	460	1.8	0.7	417	1.7	0.7	412	1.6	0.7
Pain / discomfort	529	1.5	0.6	460	1.4	0.7	418	1.4	0.9	412	1.4	0.7
Anxiety / depression	526	1.4	0.6	461	1.3	0.7	418	1.2	0.8	412	1.3	0.5
EQ-5D VAS	504	66.8	22.1	450	72.1	19.6	400	74.7	18.0	400	76.5	18.4
EQ-5D weighted	524	57.9	24.0	455	65.1	19.8	413	67.5	19.7	409	68.6	19.1
Short Form 36 (SF-36)												
Mental health	518	70.8	22.3	460	74.6	20.0	409	77.2	17.8	402	77.2	17.8
Pain	523	73.4	29.2	462	75.9	27.7	410	77.1	27.1	404	77.3	27.6
Physical	526	44.3	34.3	464	52.8	33.8	412	59.5	32.0	406	59.2	32.0
Role physical	524	23.2	37.1	464	40.5	43.4	411	51.8	43.8	405	62.2	42.6
Role emotional	519	58.0	45.0	462	70.0	40.7	410	80.3	36.0	403	83.0	33.9
Social function	522	48.6	35.1	463	65.8	32.6	411	70.2	31.0	405	75.8	28.5
Vitality	517	44.1	24.0	461	51.1	23.8	409	54.1	23.7	402	55.5	22.8
General health	520	63.7	20.3	458	64.4	19.5	411	66.4	19.9	402	66.4	20.8
PCS	511	37.0	10.3	456	39.6	11.2	404	41.3	11.6	398	42.5	11.7
MCS	511	46.1	12.1	456	49.9	11.3	404	51.8	9.8	398	52.5	10.2

Supplemental Table 3-3 Outcomes at all Times (continued)

Variable	1 month			3 months			6 months			12 months		
	n	mean	SD	n	mean	SD	n	mean	SD	n	mean	SD
Preference Based Stroke Index (PBSI)												
Walking	596	1.8	0.8	529	1.5	0.7	478	1.5	0.6	450	1.4	0.6
Stairs	595	1.9	0.8	526	1.6	0.7	475	1.5	0.6	449	1.6	0.7
Physical activities	596	2.6	0.6	528	2.4	0.7	475	2.4	0.7	446	2.3	0.7
Recreational activities	593	2.0	0.8	528	1.9	0.8	476	1.9	0.8	446	1.8	0.8
Work / activity	592	2.3	0.7	528	2.1	0.7	474	2.1	0.8	447	2.0	0.8
Driving	588	3.0	1.0	529	2.8	1.1	473	2.7	1.2	446	2.6	1.3
Memory	590	1.4	0.6	527	1.3	0.6	472	1.3	0.5	448	1.4	0.6
Speech	596	1.2	0.5	528	1.2	0.5	474	1.1	0.4	449	1.1	0.4
Coping	582	1.4	0.7	523	1.3	0.6	470	1.3	0.6	445	1.3	0.6
Self esteem	575	1.3	0.5	518	1.2	0.5	466	1.2	0.5	444	1.2	0.5
PBSI summary	549	65.0	19.9	503	71.9	18.6	456	73.5	19.0	431	74.6	19.3

Supplemental Table 3-4 Loss to Follow-up

	Lost after time i	N total	N interviewed	N not interviewed	Total lost after time i
At 1 month			618	60	
Lost by next interview:					78
Died	8				
Long-Term Care	15				
Lost to Follow-up	18				
Moved out of region	1				
Refused further participation	36				
At 3 months		600	533	67	.
Lost by next interview:					58
Died	4				
Lost to Follow-up	22				
Long-Term Care	8				
Moved out of region	2				
Refused further participation	22				
At 6 months		542	486	56	
Lost by next interview:					87
Died	8				
Lost to Follow-up	58				
Long-Term Care	5				
Refused further participation	16				
At 12 months	.	454	454	0	.

Supplemental Table 3-5 Testing for Normality in HRQL Indices

Variable	Normality
Summary Scores	
EQ5D 10 yr weights /100	Moderately non-normal
HUI3 multiattribute /100	Moderately non-normal
PBSI summary /100	Moderately non-normal
Mental health summary SF-36 /100	Moderately non-normal
Physical health summary SF-36 /100	Moderately non-normal
SIS physical summary /100	Moderately non-normal
Subscale Scores	
PBSI walking, stairs, physical activities, recreational activities, work / activity, driving, memory, coping, self esteem	Moderately non-normal and ordinal categorical
PBSI speech	Severely non-normal
EQ5D mobility, self care, usual activities, pain, anxiety / depression,	Moderately non-normal and ordinal categorical
EQ5D VAS	Moderately non-normal
HUI3 vision , hearing, speech, emotion, speech, cognition	Severely non-normal
HUI3 ambulation, dexterity, pain, self care	Moderately non-normal
SIS recovery, hand, mobility, ADL, communication, emotional, memory, social participation,	Moderately non-normal
SF-36 physical function, role physical, role emotional, social function, mental health, pain, vitality, general health	Moderately non-normal

Moderately non-normal = skewness<2, kurtosis<7, severely non-normal = skewness>2, kurtosis>7 ³³

Following Chapters

The response shift susceptible HRQL constructs of mental health, physical function, and participation are presented in individual chapters with linking chapters in between, followed by a conclusion chapter. Identification of response shift in the mental health model is presented as an example of the use of a framework to assist in response shift identification with SEM. This is presented in Chapter 5, “A Framework for Testing Response Shift in Health Related Quality of Life over Multiple Time Points – A Demonstration with Individuals Post Stroke”. The identification of response shift in the physical function model is presented in the paper in Chapter 7, “Can Response Shift Occur in Self Perceived Physical Function? Using Stroke as a Model”. The participation model is briefly discussed in Chapter 8.

Reference List

- (1) Ahmed S, Mayo N, Wood-Dauphinee S, Hanley J, Cohen S. Response Shift Influenced Estimates of Change in Health-Related Quality of Life Poststroke. *Journal of clinical epidemiology* 2004;57:561-70.
- (2) Ahmed S, Mayo N, Corbiere M, Wood-Dauphinee S, Hanley J, Cohen R. Using the Patient Generated Index to Evaluate Response Shift Post Stroke. *Quality of Life Research* 2005;14:2247-57.
- (3) Ahmed S, Mayo N, Corbiere M, Wood-Dauphinee S, Hanley J, Cohen R. Change in Quality of Life of People with Stroke Over Time: True Change or Response Shift? *Quality of Life Research* 2005;14:611-27.
- (4) Schmitt N. The Use of Analysis of Covariance Structures to Assess Beta and Gamma Change. *Multivariate Behavioural Research* 1982;17:343-58.
- (5) Oort FJ. Using Structural Equation Modeling to Detect Response Shifts and True Change. *Quality of Life Research* 2005;14:587-98.
- (6) Sprangers MA, Schwartz CE. Integrating Response Shift into Health-Related Quality of Life Research: A Theoretical Model. *Social Science and Medicine* 1999;48:1507-15.
- (7) Oort FJ, Visser M, Sprangers MA. An Application of Structural Equation Modeling to Detect Response Shifts and True Change in Quality of Life Data from Cancer Patients Undergoing Invasive Surgery. *Quality of Life Research* 2005;14:599-609.

- (8) Nolte S, Elsworth G, Sinclair A, Osborne R. Replication of a Structural Equation Modeling Approach to Detect Response Shift in Participants of Health Education Programs. *Quality of Life Research* October[suppl.], A-13. 2007.
Ref Type: Abstract
- (9) Mayo N, Scott S, Dendukuri N, Ahmed S, Wood-Dauphinee S. Identifying Response Shift Statistically at the Individual Level. *Quality of Life Research - accepted* 2008.
- (10) Borsboom G, Korfage IJ, Essink-Bot ML, Duivenvoorden H. Letter to the Editor - The Structural Equation Modeling Technique Did Not Show a Response Shift, Contrary to the Results of the Then Test and the Individualized Approaches. *Journal of clinical epidemiology* 2007;60:426-8.
- (11) Ahmed S, Mayo N. Response to Donaldson's Commentary. *Quality of Life Research* 2005;14:2357-8.
- (12) Wood-Dauphinee S. Quality of Life Post Stroke. *Brainwaves - newsletter for members of the Canadian Stroke Network* 2003;Issue 19(May):2-4.
- (13) Cote R, Hachinski V, Shurvell B, Norris J, Wolfson C. The Canadian Neurological Scale: A Preliminary Study in Acute Stroke. *Stroke* 1986;17(4):731-7.
- (14) Manitoba Centre for Health Policy. Charlson Index. *Manitoba Centre for Health Policy Concept Dictionary* 2007 February 19;Available at: URL: <http://mchp->

appserv.cpe.umanitoba.ca/viewConcept.php?conceptID=1098#a_references.

Accessed May 26, 2008.

- (15) Roccaforte W, Burke W, Bayer B, Wengel S. Validation of a Telephone Version of the Mini-Mental State Examination. *Journal of the American Geriatrics Society* 1992;40:697-702.
- (16) Wade D, Collin C. The Barthel Index: a Standard Measure of Physical Disability? *Disability Studies* 1988;10:64-7.
- (17) Jutai J, Day H. Psychosocial Impact of Assistive Devices Scale (PIADS). *Technology and Disability* 2002;14:107-11.
- (18) Burholt V, Windle G, Ferring D, Balduccim C, Fagerström C, Thissen F, Weber G, Wenger G. Reliability and Validity of the Older Americans Resources and Services (OARS) Social Resources Scale in Six European Countries. *J Gerontol B Psychol Sci Soc Sci* 2007;62(6):S371-S379.
- (19) Anderson C, Laubscher S, Burns R. Validation of the Short Form 36 (SF-36) Health Survey Questionnaire Among Stroke Patients. *Stroke* 1996;27:1812-6.
- (20) Dorman P, Waddell F, Slattery J, Dennis M, Sandercock P. Is the EuroQol a Valid Measure of Health-Related Quality of Life After Stroke? *Stroke* 1997;28:1876-82.
- (21) Mathias S, Bates M, Pasta D, Cisternas M, Feeny D, Patrick D. Use of the Health Utilities Index with Stroke Patients and Their Caregivers. *Stroke* 1997;28(10):1888-94.

- (22) Duncan P, Wallace D, Lai S, Johnson D, Embretson S, Laster L. The Stroke Impact Scale Version 2.0. *Stroke* 1999;30:2131-40.
- (23) Poissant L, Mayo N, Wood-Dauphinee S, Clarke A. The Development and Preliminary Validation of a Preference-Based Stroke Index (PBSI). *Health and Quality of Life Outcomes* 2003 September 10;1(43).
- (24) SAS Institute. *SAS/STAT User's Guide, Version 9*. Cary,NC: SAS Institute, Inc.; 2004.
- (25) Jöreskog K, Sörbom D. *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Lincolnwood, IL: Scientific Software International Inc.; 1993.
- (26) Stewart A. The Medical Outcomes Study Framework of Health Indicators. In: Stewart A, Ware J, editors. *Measuring Functioning and Well-Being*. Durham: Duke University Press; 1992. p. 12-24.
- (27) Fairclough D, Gagnon D, Papadopoulos G. Planning Analysis of Quality of Life Studies: A Case Example with Migraine Prophylaxis. *Journal of biopharmaceutical statistics* 2004;14(1):31-51.
- (28) Twisk J. Missing Data in Longitudinal Studies. *Applied Longitudinal Analysis for Epidemiology - A Practical Guide*. Cambridge: Cambridge University Press; 2003. p. 202-25.

- (29) Fairclough D, Peterson H, Chang V. Why are Missing Quality of Life Data a Problem in Clinical Trials of Cancer Therapy? *Statistics in medicine* 1998;17:667-77.
- (30) Côté R, Battista R, Wolfson C, Boucher J, Adam J, Hachinski V. The Canadian Neurological Scale: Validation and Reliability Assessment. *Neurology* 1989;39:638-43.
- (31) Collin C, Wade D, Davies S, Horne V. The Barthel Index:a Reliability Study. *International Disability Studies* 1988;10:61-3.
- (32) Enders C. Analyzing Structural Equation Models with Missing Data. In: Hancock G, Mueller R, editors. *Structural Equation Modeling - A Second Course*.Greenwich: Information Age Publishing; 2006. p. 313-42.
- (33) Finney S, DiStefano C. Non-Normal and Categorical Data in Structural Equation Modeling. In: Hancock G, Mueller R, editors. *Structural Equation Modeling: A Second Course*.Greenwich: Information Age Publishing; 2006. p. 269-314.
- (34) Bollen K. Structural Equation Models: Two Special Topics. 5 A.D.; Toronto: SPIDA, Institute for Social Research, York University; 2006.
- (35) Sprangers MA. Quality-of-life assessment in oncology. Achievements and challenges. *Acta oncologica Stockholm, Sweden* 2002;41(3):229-37.

- (36) World Health Organization. Towards a Common Language for Functioning, Disability and Health - ICF. *WHO website* 2002; Available at: URL: www3.who.int/icf/beginners/bg.pdf. Accessed April 12, 2006.
- (37) Donaldson G. Structural Equation Models for Quality of Life Response Shifts: Promises and Pitfalls. *Quality of Life Research* 2005;14:2345-51.
- (38) Hancock G. Structural Equation Modeling - A One Day Introductory Workshop. 2005.
- (39) Kline R. *Principles and Practice of Structural Equation Modeling*. 2 ed. New York: The Guilford Press; 2005.
- (40) Duncan P, Reker D, Horner R, Samsa G, Hoenig H, LaClair B, Dudley T. Performance of a Mail-Administered Version of a Stroke-Specific Outcome Measure, the Stroke Impact Scale. *Clinical Rehabilitation* 2002;16:493-505.
- (41) Drummond M, O'Brien BJ, Stoddart G, Torrance G. Cost Utility Analysis. *Methods for the Economic Evaluation of Health Care Programmes*. 2nd ed. Oxford: Oxford University Press; 1997. p. 139-203.
- (42) Grootendorst P, Feeny D, Furlong W. Health Utilities Index Mark 3: Evidence of Construct Validity for Stroke and Arthritis in a Population Health Survey. *Medical Care* 2000;38(3):290-9.
- (43) Duncan P, Lai S, Tyler D, Perera S, Reker D, Studenski S. Evaluation of Proxy Responses to the Stroke Impact Scale. *Stroke* 2002;33:2593-9.

- (44) Pierre U, Wood-Dauphinee S, Korner-Bitensky GD, Hanley J. Proxy Use of the Canadian SF-36 in Rating Health Status of the Disabled Elderly. *Journal of Clinical Epidemiology* 1998;51(11):983-90.
- (45) Dorman P, Waddell F, Slattery J, Dennis M, Sandercock P. Are Proxy Assessments of Health Status After Stroke With the EuroQol Questionnaire Feasible, Accurate, and Unbiased? *Stroke* 1997;28:1883-7.
- (46) Dorman P, Slattery J, Farrell B, Dennis M, Sandercock P. Qualitative Comparison of the Reliability of Health Status Assessments with the EuroQol and the SF-36 Questionnaires after Stroke. *Stroke* 1998;29:63-8.
- (47) Finch E, Brooks D, Stratford P, Mayo N. *Physical Rehabilitation Outcome Measures*. 2nd ed. Hamilton: BC Decker Inc; 2002.

Chapter 4: Linking Literature and Methods to a Statistical Framework for Examining Response Shift

The literature is clear that QOL and HRQL are important for understanding and quantifying recovery following stroke, but using composite measures encompassing several constructs may mask response shift. These composite measures incorporate several separate constructs which may themselves be subject to response shift in differing directions and times post-stroke. The literature is also clear that statistical methods for identifying response shift need to be developed as it is often impractical to incorporate direct questioning of respondents about response shift or retrospective ratings of previously elicited values into a study design.

Chapter 3 outlined the statistical approaches undertaken to further advance the understanding of response shift in a stroke population. The particular focus is to identify response shift susceptible constructs from global measures of HRQL. Available for these analyses was a rich data source comprising many HRQL indices. The ICF model was used to frame the constructs for study, some of which were hypothesized to show response shift while others were not.

The following chapter outlines a framework for identifying response shift occurring over more than two time points using SEM. The manuscript that follows has been invited for submission to a special issue on Response Shift in the Journal of Clinical Epidemiology.

It is important set guidelines for analyses when there are multiple time points because there are a large number of possible localizations for response shift when considering

timing, type, and specific measured variable. This complexity has a direct impact on the specification of the Type I error, as multiple testing is carried out. The framework also guides choices depending on the intent of the response shift analysis. If the study intent is it to inform future research, then an exploratory analytical approach would be appropriate; if the intent is to confirm past research, then a theory-driven or confirmatory analytical approach would be chosen.

To illustrate the application of this framework, we used the construct of mental health. Under the ICF model, many of the items included in mental health indices refer to impairments of emotion and cognition. The results of the analyses illustrated in the manuscript in Chapter 5 showed that response shift was identified in the mental health construct primarily at the 12 month period. Methodologically, the value of the framework in guiding the analyses was illustrated.

Chapter 5: Response Shift was Identified in Health Related Quality of Life Using a Framework Developed for Multiple Time Points: A Demonstration with Individuals Post Stroke

Invited for submission to the special issue on response shift in the Journal of Clinical Epidemiology, “Enhancing the Application of Health-Related Quality of Life (HRQL) Evaluations for Guiding Clinical Care: Advancements in Response Shift Research”.

Submitted August 13, 2008.

Ruth Barclay-Goddard BMR(PT), MHSc¹

Lisa Lix PhD²

Robert Tate PhD³

Leah Weinberg PhD¹

Nancy E. Mayo, BSc(PT), MSc, PhD⁴

1. Department of Physical Therapy, School of Medical Rehabilitation, University of Manitoba, R106-771 McDermot Ave. Winnipeg, Manitoba, Canada R3E 0T6
barclayg@cc.umanitoba.ca Phone (204)977-5637, fax (204)789-3927

2. School of Public Health, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

3. Department of Community Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada

4. Department of Medicine and School of Physical and Occupational Therapy, McGill University; Division of Clinical Epidemiology; Division of Geriatrics, McGill University Health Center, Royal Victoria Hospital Site, R4.29, 687 Pine Ave W, Montreal, QC, H3A 1A1 nancy.mayo@mcgill.ca Phone (514) 934 1934 36922, fax (514) 843 1493

Abstract

Objective: To develop and apply a framework that uses structural equation modeling (SEM) to identify response shift in data with greater than two time points. **Study Design**

and Setting: The framework addresses key issues that arise when analyzing data with multiple time points using a model-based approach to test for response shift: model validation, correction for multiple testing, and adoption of an exploratory or theory-driven approach to identify the type and timing of response shift. Data from an observational study of 678 individuals at 1, 3, 6, and 12 months post stroke are used to demonstrate the application of the framework to a model for mental health. **Results:**

Uniform and non-uniform recalibration was identified at 6 and 12 months post stroke.

Conclusion: Studies that identify the type and timing of response shift in certain client populations may be useful for planning the timing of treatment and the methods to measure response shift clinically. Validation of the model as well as adjusting for the effects of multiple testing increases confidence in the mental health model and the resulting identification of response shift.

Word count:

abstract = 181

manuscript = 3280

Key words:

health related quality of life

longitudinal

mental health

response shift

stroke

structural equation modeling

Acknowledgements:

The data source from the original longitudinal study, “Understanding Quality of Life Post-Stroke: A Study of Individuals and their Caregivers” was funded by the Canadian Stroke Network.

What is new?

- The framework was successful in addressing issues of model validation, multiple testing, and the purpose of a study in the identification of response shift in greater than two occasions with current structural equation modeling methodology.
- Recalibration response shift was identified primarily at 12 months post stroke in mental health.
- This has potential implications for the clinical assessment and timing of the application of treatment aimed at encouraging response shift.

Introduction

The analysis of longitudinal health related quality of life (HRQL) is challenged by the potential for persons to experience ‘response shift’ over time. Response shift is defined as “a change in the meaning of one’s self-evaluation of a target construct as a result of: (a) a change in the respondent’s internal standards of measurement (i.e., scale recalibration); (b) a change in the respondent’s values (i.e., the importance of component domains constituting the target construct) or (c) a redefinition of the target construct (i.e., reconceptualization)”[1].

Response shift is a potential explanation when an individual with a health event or chronic condition describes HRQL consistent with the HRQL of healthy individuals; this is unexpected but not uncommon. Among individuals who have survived a stroke, HRQL ratings can be similar to the ratings given by peers despite decreased physical function and walking ability [2]; individuals six months post stroke have demonstrated similar quality of life to that of controls [3]. With response shift, the concept of HRQL changes over time but tests can not be compared longitudinally due to changes in internal standards, values, and/or concepts. True change may be over or under-estimated when response shift is present, leading to biased estimates of the magnitude of change.

Many methods have been used to detect response shift in longitudinal data, however, the focus of this manuscript is on structural equation modeling (SEM). SEM is a multivariate regression technique for the analysis of data containing measurement error [4-7].

Research describing SEM to test for the presence of response shift is relatively recent and

has only been described for two measurement occasions (i.e., pre- and post- data); there is an opportunity to extend this technique to multiple time points.

The manuscript begins by introducing the use of SEM to test for response shift at two measurement occasions. Data from the observational study “Understanding Quality of Life Post-Stroke: A Study of Individuals and their Caregivers” [8] is used to demonstrate the proposed framework. Opportunities for SEM / response shift research and clinical applications follow.

Structural Equation Modeling and Response Shift

SEM is a family of techniques for the analysis of latent variables. Latent variables are variables that can not be measured directly, but are inferred by and associated with observed measured variables [9]. These relationships are referred to as paths or factor loadings, and are comparable to regression coefficients in multiple regression analysis. Changes in parameter estimates (paths, error variances, and intercepts) over time are indicative of response shift.

Oort proposed a method using SEM to identify the presence of response shift (reconceptualization, reprioritization, uniform and non-uniform recalibration) in longitudinal data [4;5]. The method was demonstrated with two time points and detected response shift among cancer patients receiving surgery [5]. Oort proposed that changes in factor loading patterns over time indicate reconceptualization [4]. This is demonstrated when a measured variable is associated with one latent variable, but at a later point in time is associated with a different latent variable. Changes in the magnitude of factor loadings represents reprioritization [4]; the association between a measured variable and

a latent variable may become stronger or weaker over time, reflecting changes in values or priorities. Uniform recalibration occurs when the entire measurement scale is recalibrated; non-uniform recalibration occurs when a portion of the scale is recalibrated [4]. Uniform recalibration is suggested by differences between intercepts and non-uniform recalibration by differences between error variances across occasions [4].

Testing for response shift with multiple measurement occasions is more complex than with two occasions. In both instances the researcher must decide whether tests of response shift should focus on specific types of response shift, on specific measured variables, and with multiple time points, which specific time points to consider. This leads to multiple testing and increased chances of a Type I error (concluding there is a difference when there is not). In addition, the focus of a study may be theory-driven, confirming results of previous studies or it can be exploratory to identify response shift, informing future studies. This paper addresses these issues within a framework to guide future response shift research using SEM with multiple occasions.

Development of a Framework to Test Response Shift

Readers are encouraged to review Oort's paper for detailed explanation of the methodology [4]; the framework presented here builds on that SEM methodology. The framework is presented in Figure 1. In **Step 1**, response shift presence is tested using an overall likelihood ratio test (chi-square (χ^2) difference test). This corresponds to Oort's step 2 [4]. Two models are compared, the free model in which parameter estimates (paths, error variances, and intercepts) are all freely estimated, and the fully constrained model in which all of the paths, error variances, and intercepts are made to be equal

across time [4]. The χ^2 difference test evaluates the change in overall fit of the free and fully constrained (no response shift) models [10]. When the χ^2 difference test is statistically significant, the null hypothesis that the two models fit equally well is rejected, providing evidence response shift is present in the data.

In **Step 2**, validation of the model(s) in Step 1 is completed. Validation is the process of determining that a model is good [11]. A model is not proven by having a good fit; a model could be a good model for the population or it could fit the sample only [10].

Three types of validation have been described: apparent accuracy, internal validation, and external validation [12]. Apparent accuracy involves the evaluation of different methods when estimating a model with a particular sample (i.e., a sensitivity analysis). This is demonstrated by comparing a method that assumes normality, and one that does not.

Internal validation, which examines the performance of a model in the same population from which the sample was drawn, relies on such techniques as data splitting (using part of the sample to develop the model, and the other part to confirm the model), cross-validation (repeated data splitting), and bootstrapping (a technique of resampling from the population to obtain estimates of summary statistics) [9;11-13]. External validation is considered the most rigorous by applying the model to data from a different population [12].

Step 3 represents the decision that a researcher makes when designing the study and is reflected in the framework as a choice of paths (exploratory or theory-driven). An exploratory approach is chosen when there is little or no evidence about the potential sources of response shift within a population. The purpose of the analysis is to identify

potential sources of response shift to inform the literature for future research. Even though the process is exploratory, we do not suggest that only empirical data (i.e., modification indices) be used, but that possible sources of response shift tested are ‘theory supported’, given current response shift and population knowledge. A modification index is a χ^2 that estimates the amount the overall model χ^2 would decrease if changes are made to the model.

In the exploratory approach, modification indices (as described by Oort) will guide the initial search for testing of response shift identification; factor loadings, intercepts, and error variances at various times and measured variables will be evaluated [4]. If response shift at a particular time does not make sense based on knowledge of the condition (e.g. at baseline), that time will not be tested, even if modification indices suggest that time.

In comparison, a theory-driven approach is chosen when there is sufficient evidence (previous studies or theory) which supports analyses of particular timing, type, or measured variables that contribute to the rejection of the global null hypothesis of no response shift. The purpose of the theory-driven approach is to confirm current evidence of response shift in the population studied. For example, if a previous study suggested response shift at six months, the search for response shift may concentrate on that time-frame, and modification indices will guide the researcher to further focus on removing constraints that suggest particular types of response shift.

Step 4 is the application of the exploratory or theory-driven approach. Constraints are removed one at a time from the constrained model to determine the sources of response shift [4]. With the removal of each constraint (e.g., time two intercept for variable x), the

new partially constrained model is compared to the fully constrained model using the χ^2 difference test. This corresponds to Oort's step 3 [4]. A statistically significant result suggests response shift related to the freed constraint (i.e., uniform recalibration at time two in variable x).

Step 5 addresses the possibility of making a Type I error with multiple testing; this is the familywise Type I error rate [14]. A family includes all of the inferences that are made in a study; the familywise Type I error rate can be controlled by adjusting the p values [14]. The following example uses the Bonferroni correction, which is widely used because it is simple to apply. The α significance level is adjusted down by dividing the α significance level (typically 0.05) by the number of hypotheses tested [14-16]. Modified Bonferroni procedures, as described by Hochberg and by Holm, could be used. The Bonferroni can be too conservative, leading to higher probability of Type II error [17]. Both methods evaluate p values in a sequential order with varying α significance level adjustments for each p value [17].

Application of Framework

Description of Example Dataset

Data used in this analysis are from 677 persons with stroke collected in three Canadian cities at 1, 3, 6, and 12 months post-stroke. The HRQL indices used to measure mental health are described in Table 5-1. A description of the study participants is provided in Table 5-2. Mean scores for all measured variables of mental health across time are in

Table 5-3. Complete data at all four time points was available for 344 (50.4%) participants; 67% of participants completed the study.

Content of the mental health model was based on the Medical Outcomes Study (MOS) Framework of Health Indicators [23]. The MOS describes mental health as: cognitive functions and psychological distress / wellbeing [23]. The two latent variables in the following model are described as ‘cognitive function’ and ‘psychological function’.

Table 5-1 also lists the measured variables associated with each latent variable.

Preparatory work to develop the model

Data preparation, data screening, and imputation used SAS 9.1 software [18]. LISREL 8.7 SIMPLIS and PRELIS software [19] were used to construct and test the models.

Prior to testing for response shift, an appropriate model, was identified [4]. The fit was assessed in numerous ways. When model χ^2 increases, the fit of a model become worse; it tests the difference between the observed model and a model that has a perfect population fit [10]. A non-significant χ^2 means there is little difference between the models, suggesting a good fit. Models with large sample sizes, however, can often be rejected because the model χ^2 is affected by sample size [20]. The normed χ^2 statistic, the ratio of the χ^2 statistic to its degrees of freedom (df), decreases the effect of the sample size on the χ^2 ; values up to 5.0 suggest a reasonable model fit [10]. The root mean square error of approximation (RMSEA) measures the lack of fit in a model compared to the population. A value no greater than 0.05 is indicative of close model fit, a value between 0.05 and 0.08 is indicative of reasonable fit, and a value of 0.10 or greater is indicative of poor model fit [21]. The comparative fit index (CFI) has a range of 0-1.0; a value of >0.90

suggests a reasonably good fit [22]. Values of less than 0.10 are acceptable for the standardized root mean square residual (SRMR) [10]. The longitudinal model was based on the model that was developed for the data at the baseline (one month) measurement occasion. This model is presented in Figure 2. It had a close fit with a RMSEA of 0.048 (90% confidence interval 0.033 - 0.064). The χ^2 statistic was 60.58, with 25 df, p value of 0.000, and a normed χ^2 of 2.42. The longitudinal model (1, 3, 6, and 12 months) had a close fit with RMSEA of 0.033 (0.030 -; 0.037), χ^2 890.20, 508 df, p value of 0.00, and a normed χ^2 of 1.75. Model fit focused on the RMSEA and normed χ^2 due to the common poor fit of χ^2 in large samples. Only χ^2 and RMSEA fit statistics were available with full information maximum likelihood (FIML) estimation in LISREL. FIML accounts for missing data. With FIML, missing data are not imputed; parameter estimates are estimated from raw data with iterative computer algorithms [24]. The parameter estimates that FIML produces are unbiased under the assumption of missing at random; missing data was determined to be missing at random [25]. FIML also assumes normality, however, the measured variables were moderately non-normal [24]. Robust maximum likelihood estimation (MLE) is robust with non-normal data. Five datasets were imputed and analyzed robust MLE, which provides a Satorra-Bentler (SB) χ^2 statistic; the SB χ^2 performs better than the χ^2 from non-robust methods when data are continuous or ordered categorical non-normal [26].

Step 1 – Identifying the Presence of Response Shift

The χ^2 difference statistic for the longitudinal free and fully constrained models was 403.45 with 72 df ($p < 0.0001$), suggesting the presence of response shift. The covariance and means matrices produced by FIML are available on request from the authors.

Step 2 - Validation of the model

In this example, we chose an apparent validation (sensitivity analysis) which compared two methods of parameter estimation as an example of validation. Model fit was compared across the five imputed datasets using robust MLE and FIML. All robust longitudinal models had a reasonable fit (RMSEA 0.057 – 0.060, CFI = 0.98, and SRMR 0.061 – 0.066). [See Appendix 5-1 for thesis only.] Both methods of parameter estimation appeared appropriate to use with the data. The remainder of the analysis is based on FIML estimation.

Step 3 Identify Localized Sources(s) of Response Shift

A choice is made between the exploratory or theory-driven approach, when designing a study. Both methods, however, are demonstrated below.

Step 4 - Exploratory Approach – Test Sources that are “Theory Supported”

Using the modification indices from the fully constrained longitudinal model and knowledge of the mental health model, eleven constraints were removed one at a time. Each new model represented a potential source of response shift. A Bonferroni correction was used to control the familywise Type I error rate (**Step 5**). We defined the ‘family’ as

all the constraints freed in the model. Using a significance level of $\alpha = 0.05 / 11 = .0045$), uniform and non-uniform recalibration were identified at 6 and 12 months in six locations. Table 5-4 shows the parameter changes over time in the six response shift susceptible variables. All response shift occurred within the psychological function latent variable. Applying the Holm and Hochberg [17] procedures, the same results were obtained.

Step 4 – Theory-driven Approach – ‘a priori’

Previously, reconceptualization, reprioritization, and recalibration post stroke have been suggested between 6 weeks and 6 months [27;28]. We therefore chose to focus on three months, which is in the same time range. Response shift was not identified, however, in the same time period with an alternate SEM methodology [29]. Our a-priori hypothesis was that there would be response shift identified at 3 months. Non-uniform recalibration of the PBSI memory subscale was identified at 3 months by freeing the error variance. The χ^2 difference test was statistically significant ($\chi^2 = 5.70$, 1 df, $p = 0.017$). Correcting for multiple testing was not required because only a single test was conducted.

Discussion

The framework was developed as a guide for researchers who use SEM to test for response shift. The framework builds on the work of Oort, but focuses specifically on issues that arise when longitudinal data are collected at more than two measurement occasions. The issues addressed in the framework include: model validation, type of model assessment (exploratory versus theory-driven), and multiple testing.

With the exploratory approach, recalibration response shift (uniform and non-uniform) was identified for the mental health construct of HRQL in individuals post stroke at 6 and 12 months. Uniform recalibration was identified at 12 months in SF-36 role emotional and PBSI self esteem. Non-uniform recalibration was identified at 6 months in SF-36 mental health and 12 months in EQ5D anxiety/depression, SF-36 mental health and SIS emotional.

In comparison, response shift in HRQL was not identified between 6 weeks and 6 months using another SEM method previously [7;29]. Differences in results could be from the method used, the timeframe, or the model. The current study used a later timeframe, and different models were used between the two studies. This study developed a mental health model using numerous HRQL indices, while the earlier model was based only on the SF-36.

With the theory-driven approach, response shift was examined at 3 months. Non-uniform recalibration was identified in the PBSI memory subscale. Similarly, previous research suggested recalibration post stroke using a different method of assessing response shift, the *then* test [27].

A choice between the exploratory and theory-driven approach needs to be made, based on available evidence. In this demonstration, response shift was not identified at 3 months in the exploratory approach due to correction for familywise Type I error. Results between the approaches may vary, due to the number of models tested. There is still much to be done to understand the phenomenon of response shift post stroke as well as for individuals with other health conditions. Until more is known, the utilization of an

exploratory approach, as outlined in this framework, will lead to the identification of potential sources of response shift, to inform future research.

There are some limitations to the framework. Validation of the longitudinal model was demonstrated, however, there is also opportunity to assess validation through identifying response shift susceptible variables with both FIML and robust MLE. One could argue that validation of the cross-sectional model should also occur. Power and sample size in large longitudinal models need consideration. There were more than 100 parameters estimated in the unconstrained model. It has been suggested that the number of cases to free parameters estimated be at least ten to one [20]; the sample might therefore be considered small. Sample size requirements increase when the model is complex due to the inclusion of multiple time points, latent variables or measured variables.

Future research could utilize the framework to compare the SEM methods described by Oort and Schmitt with the same sample [4;7]. Application of the framework to other HRQL constructs with the data used in this study is ongoing; response shift has been suggested in participation and physical function models. In this demonstration, response shift was identified at 6 and 12 months post stroke; the possibility that the catalyst for response shift is getting back into life at home and in the community requires study.

Conclusions

We developed a framework to identify response shift using SEM with more than two time points. Validating the model increases confidence in the model and controlling for Type I error strengthens acceptance of the results.

Improved HRQL and response shift are often goals of rehabilitation [30-33]. Response shift should therefore be measured in research studies and clinically; if response shift is present, as in this demonstration, the calculation of changes over time may be inaccurate.

The framework may help guide clinically focused research, such as identifying the timing of response shift for the application of treatment aimed at encouraging response shift.

This could assist in the development of clinical practice guidelines for the measurement of response shift and HRQL over time, such as when to measure for response shift and what type of response shift to evaluate in the clinic.

Figure 5-1 Identifying Response Shift with Multiple Time Points

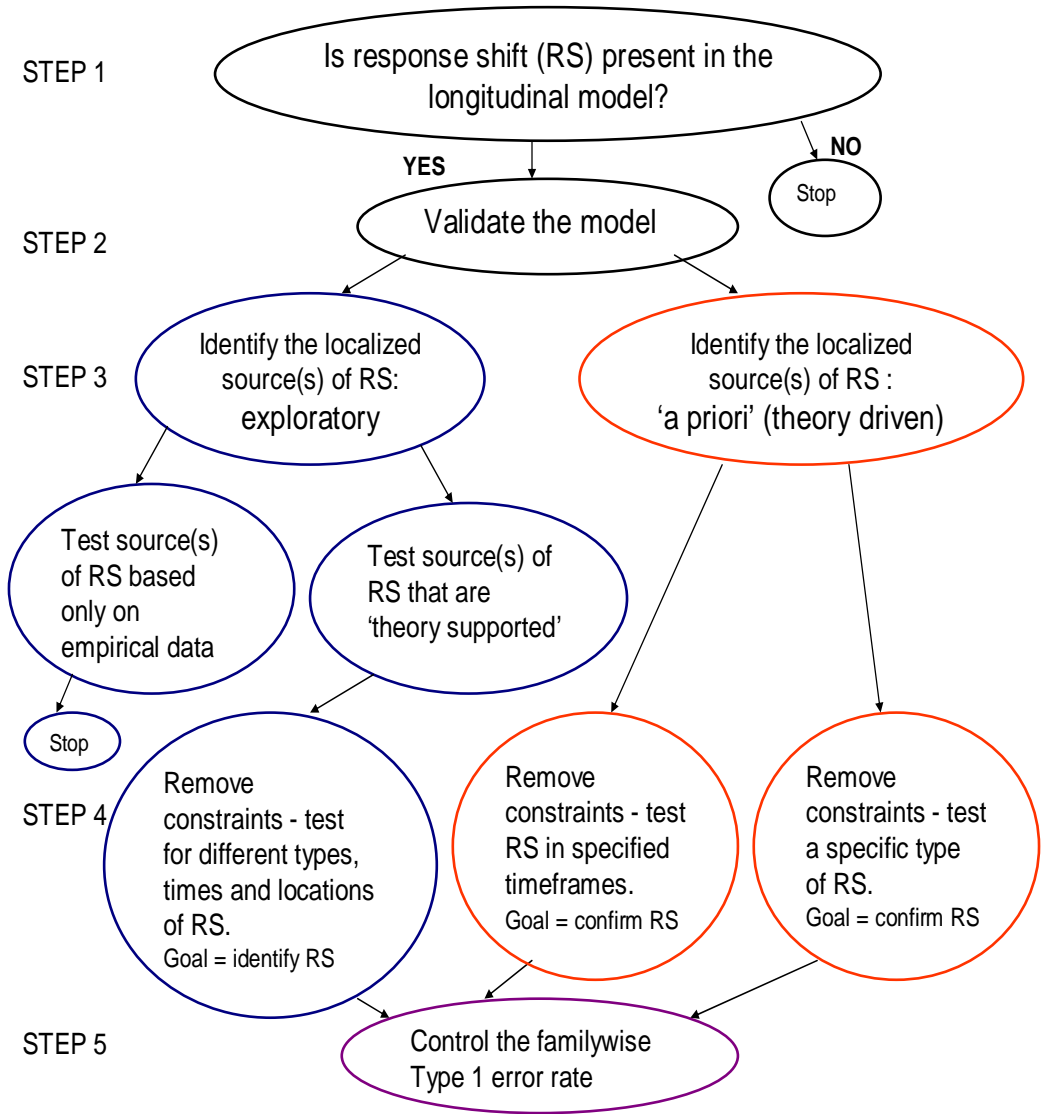
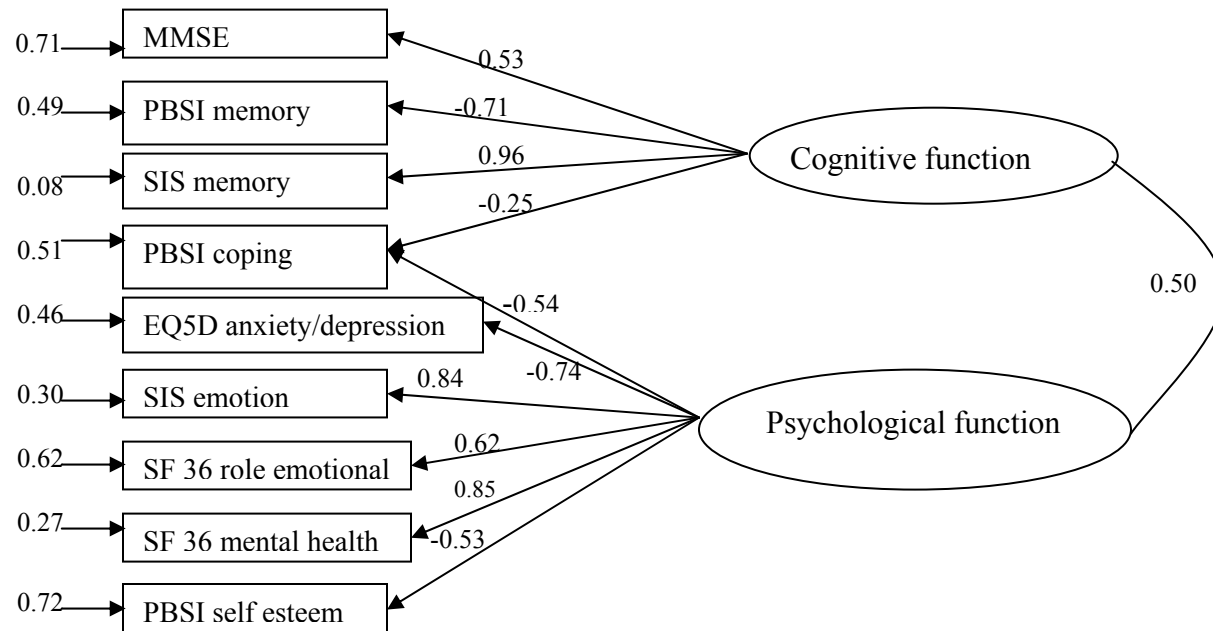


Figure 5-2 Mental Health Model – 1 month



Standardized solution: $\chi^2 = 60.58$, $df=25$, p value = 0.00009, RMSEA = 0.048
 Negative correlations reflect the measures PBSI and the EQ5D, where the highest score is the worse health condition.
 MMSE = Mini-Mental State Exam, SIS = Stroke Impact Scale, EQ5D = Euroqol, SF-36 = Short Form 36, PBSI = Preference Based Stroke Index

Table 5-1 HRQL Indices and Variables Utilized in the Mental Health Model

Description	Item number and content	Scale	Cognitive function	Psychological function
Mini Mental State Exam (MMSE) telephone version (34) Screen for cognitive impairment	telephone version is scored out of 22 (no reading, writing, copying)	rescaled / 100	total score	-
Short Form-36 (SF-36) (35;36) Generic health profile Validity and reliability evaluated in clients with stroke.	8 domains: physical functioning, role limitations due to physical problems, bodily pain, vitality, social functioning, role limitations due to emotional problems, mental health, and general health	domain scores and summary scores of mental and physical health /100	-	mental health role emotional
EuroQuol (EQ5D) (36-38) Generic utility index Validity and reliability evaluated in stroke	5 domains: mobility, self care, usual activities, pain / discomfort, anxiety / depression plus visual analogue scale (VAS) of self perceived health status	single domain scores or multiattribute utility scores. Scaling in each domain = 1-3	-	anxiety/ depression
Health Utilities Index (HUI) version 2/3 (38-40) Generic utility index Reliability in a general population and validity in stroke have been measured	9 domains: emotion, cognition (memory, thinking), self-care, pain / discomfort, vision, hearing, speech, ambulation, and dexterity	single domain scores or multiattribute utility scores = 0-1	cognition*	emotional*
The Stroke Impact Scale (SIS) (41;42) Stroke specific Validity and reliability have been assessed	8 domains: strength, hand function, mobility, activities of daily living and instrumental activities of daily living, emotion, memory, communication, and social participation, plus global perception of recovery (VAS)	domains / 100	memory	emotion
Preference Based Stroke Index (PBSI) (43) Preference based stroke specific Content and construct validity have been tested	10 items: self esteem, coping, speech, memory, driving, work, recreational activity, physical activity / sports, stairs, and walking.	Scaling in each item = 1-3 for variables in study.	memory coping +	self esteem coping +

*extremely non-normal, therefore, not used in model, + associated with both cognitive and psychological function

Table 5-2 Description of Study Participants

Characteristic	N	Value
Age (mean, SD)	676	67.3 years (14.8)
Gender (% female)	677	45.1
Side of lesion (% right, %left, % bilateral)	654	49.5, 44.7, 5.5
1 st stroke (%)	657	93.8
Function (Barthel/100) at 1 month (mean, SD) (44;45)	668	74.5 (25.9)

Table 5-3 Summary Statistics for Mental Health Variables at Four Measurement Occasions Post Stroke

Variable	1 month			3 months			6 months			12 months		
	n	Mean	SD	n	Mean	SD	n	Mean	SD	N	Mean	SD
Mini Mental State Exam – telephone version												
Adjusted / 100	502	88.9	14.8	434	91.3	12.0	399	93.1	10.3	394	93.7	10.1
Stroke Impact Scale												
Memory subscale	609	83.7	21.7	530	86.4	19.2	477	87.5	18.1	444	88.1	18.3
Emotion subscale	607	77.9	18.0	522	79.8	17.1	473	81.4	16.2	439	81.7	16.5
EuroQol / EQ5D												
Anxiety / depression	526	1.4	0.6	461	1.3	0.7	418	1.2	0.8	412	1.3	0.5
Short Form -36												
mental health	518	70.8	22.3	460	74.6	20.0	409	77.2	17.8	402	77.2	17.8
role emotional	519	58.0	45.0	462	70.0	40.7	410	80.3	36.0	403	83.0	33.9
Preference Based Stroke Index												
Memory	590	1.4	0.6	527	1.3	0.6	472	1.3	0.5	448	1.4	0.6
Coping	582	1.4	0.7	523	1.3	0.6	470	1.3	0.6	445	1.3	0.6
self esteem	575	1.3	0.5	518	1.2	0.5	466	1.2	0.5	444	1.2	0.5
Note: EQ-5D and PBSI – lower score = higher functioning; MMSE, SIS, and SF-36 – higher score = higher functioning												

Table 5-4 Localization of Response Shift: Changes in Non-Standardized Parameter Estimates

Constraint removed	1 month	3 months	6 months	12 months	1df χ^2 difference	p value*
<i>Uniform Recalibration</i>		<i>Intercepts</i>				
SF-36 role emotional	61.94	61.94	61.94	69.78	19.75	<0.0001
PBSI self esteem	1.30	1.30	1.30	1.36	9.87	0.0017
<i>Non-uniform Recalibration</i>		<i>Error variances</i>				
EQ5D anxiety/depression	0.14	0.14	0.14	0.10	15.00	0.0001
SF-36 mental health	104.32	104.32	62.63	104.32	15.81	<0.0001
SIS emotional	94.36	94.36	94.36	62.44	15.43	<0.0001
SF-36 mental health	103.41	103.41	103.41	67.96	10.47	0.0012

* α significance level $p \leq 0.0045$, only statistically significant models demonstrated

Response shift demonstrated in those indices with a parameter estimate that is different at a particular time-frame (bolded). SIS = Stroke Impact Scale, PBSI = Preference Based Stroke Index, EQ5D = EuroQol, SF-36 = Short Form 36

Appendix 5-1 Mental Health Model Validity - Sensitivity Analysis

Method	χ^2	df	Normed χ^2	S-B χ^2	Normed S-B χ^2	RMSEA (90% CI)	CFI	SRMR
FIML	890.20	508	1.75	---	---	0.033 (0.030; 0.037)	---	---
Imputed data sets with robust maximum likelihood estimation								
#1	2012.90	508	3.96	1646.10	3.24	0.058 (0.054; 0.061)	0.98	0.064
#2	2090.85	508	4.12	1706.59	3.35	0.059 (0.056; 0.062)	0.98	0.061
#3	2001.40	508	3.94	1637.86	3.22	0.057 (0.054; 0.060)	0.98	0.064
#4	2025.40	508	3.99	1666.26	3.28	0.058 (0.055; 0.061)	0.98	0.064
#5	2143.50	508	4.22	1745.80	3.44	0.060 (0.057; 0.063)	0.98	0.066
Mean (SD)	2054.81 (60.50)	508	4.04	1680.52 (45.14)	3.31	0.058 (0.001)	0.98	0.064 (0.002)

Normed χ^2 (χ^2/df) - values up to 5 indicate reasonable fit

RMSEA \leq 0.05 close fit; 0.05 to 0.08 reasonable fit; \geq 0.10 poor fit

CFI $>$ 0.90 reasonably good fit, SRMR $<$ 0.10 is favourable (11)

References

- (1) Schwartz CE, Sprangers MA. Methodological Approaches for Assessing Response Shift in Longitudinal Health-Related Quality of Life Research. *Social Science and Medicine* 1999;48:1531-48.
- (2) Mayo N, Wood-Dauphinee S, Ahmed S, Gordon C, Higgins J, McEwen S, Salbach N. Disablement Following Stroke. *Disability and Rehabilitation* 1999;21(5/6):258-68.
- (3) Mayo N, Wood-Dauphinee S, Cote R, Durcan L, Carlton J. Activity, Participation, and Quality of Life 6 Months Poststroke. *Archives of Physical Medicine and Rehabilitation* 2002;83:1035-42.
- (4) Oort FJ. Using Structural Equation Modeling to Detect Response Shifts and True Change. *Quality of Life Research* 2005;14:587-98.
- (5) Oort FJ, Visser M, Sprangers MA. An Application of Structural Equation Modeling to Detect Response Shifts and True Change in Quality of Life Data from Cancer Patients Undergoing Invasive Surgery. *Quality of Life Research* 2005;14:599-609.

- (6) Visser M, Oort F, Sprangers M. Methods to Detect Response Shift in Quality of Life Data: A Convergent Validity Study. *Quality of Life Research* 2005;14:629-39.
- (7) Schmitt N. The Use of Analysis of Covariance Structures to Assess Beta and Gamma Change. *Multivariate Behavioural Research* 1982;17:343-58.
- (8) Wood-Dauphinee S. Quality of Life Post Stroke. *Brainwaves - newsletter for members of the Canadian Stroke Network* 2003;Issue 19(May):2-4.
- (9) Hoyle R. *Structural Equation modelling - Concepts, Ideas, and Applications*. Thousand Oaks: SAGE Publications; 1995.
- (10) Kline R. *Principles and Practice of Structural Equation Modeling*. 2 ed. New York: The Guilford Press; 2005.
- (11) Last J. *A Dictionary of Epidemiology*. 4th ed. Oxford: Oxford University Press; 2001.
- (12) Harrell F, Lee K, Mark D. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine* 1996;15(4):361-87.

- (13) Schumacker R, Lomax R. A Beginner's Guide to Structural Equation Modeling. 2nd ed. London: Lawrence Erlbaum; 2004.
- (14) Ludbrook J. Statistics in Physiology and Pharmacology: A Slow and Erratic Learning Curve. *Clinical and Experimental Pharmacology and Physiology* 2001;28:488-92.
- (15) Chen J, Lin K, Huque M, Arani R. Weighted p-Value Adjustments for Animal Carcinogenicity Trend Test. *Biometrics* 2000;56:586-92.
- (16) Hassard T. Understanding Biostatistics. St. Louis: Mosby Year Book; 1991.
- (17) Hochberg Y, Benjamini Y. More Powerful Procedures for Multiple Significance Testing. *Statistics in medicine* 1990;9:811-8.
- (18) SAS Institute. SAS/STAT User's Guide, Version 9. Cary,NC: SAS Institute, Inc.; 2004.
- (19) Jöreskog K, Sörbom D. LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language. Lincolnwood, Il: Scientific Software International Inc.; 1993.

- (20) Hays R, Revicki D, Coyne K. Application of Structural Equation Modeling to Health Outcomes Research. *Evaluation and the Health Professions* 2005;28(3):295-309.
- (21) Browne M, Cudeck R. Alternative Ways of Assessing Model Fit. In: Bollen K, Long J, editors. *Testing Structural Equation Models*. London: Sage Publications; 1993. p. 136-62.
- (22) Jaccard J, Wan C. *LISREL Approaches to Interaction Effects in Multiple Regression*. Thousand Oaks: Sage Publications; 1996.
- (23) Stewart A. The Medical Outcomes Study Framework of Health Indicators. In: Stewart A, Ware J, editors. *Measuring Functioning and Well-Being*. Durham: Duke University Press; 1992. p. 12-24.
- (24) Enders C. Analyzing Structural Equation Models with Missing Data. In: Hancock G, Mueller R, editors. *Structural Equation Modeling - A Second Course*. Greenwich: Information Age Publishing; 2006. p. 313-42.
- (25) Twisk J. Missing Data in Longitudinal Studies. *Applied Longitudinal Analysis for Epidemiology - A Practical Guide*. Cambridge: Cambridge University Press; 2003. p. 202-25.

- (26) Finney S, DiStefano C. Non-Normal and Categorical Data in Structural Equation Modeling. In: Hancock G, Mueller R, editors. Structural Equation Modeling: A Second Course. Greenwich: Information Age Publishing; 2006. p. 269-314.
- (27) Ahmed S, Mayo N, Wood-Dauphinee S, Hanley J, Cohen S. Response Shift Influenced Estimates of Change in Health-Related Quality of Life Poststroke. *Journal of clinical epidemiology* 2004;57:561-70.
- (28) Ahmed S, Mayo N, Corbiere M, Wood-Dauphinee S, Hanley J, Cohen R. Using the Patient Generated Index to Evaluate Response Shift Post Stroke. *Quality of Life Research* 2005;14:2247-57.
- (29) Ahmed S, Mayo N, Corbiere M, Wood-Dauphinee S, Hanley J, Cohen R. Change in Quality of Life of People with Stroke Over Time: True Change or Response Shift? *Quality of Life Research* 2005;14:611-27.
- (30) Osborne R, Hawkins M, Sprangers M. Change of Perspective: A Measureable and Desired Outcome of Chronic Disease Self-Management Intervention Programs That Violates the Premise of Preintervention / Postintervention Assessment. *Arthritis and Rheumatism* 2006;55(3):458-65.
- (31) Wilson IB. Clinical Understanding and Clinical Implications of Response Shift. *Social Science and Medicine* 1999;48:1577-88.

- (32) Schwartz C, Andresen EM, Nosek M, Krahn G, RRTC Expert Panel on Health Status Measurement. Response Shift Theory: Important Implications for Measuring Quality of Life in People With Disability. *Archives of Physical Medicine and Rehabilitation* 2007;88:529-36.
- (33) Rochette A, Korner-Bitensky N, Levasseur M. 'Optimal' Participation: A Reflective Look. *Disability and Rehabilitation* 2006;28(19):1231-5.
- (34) Roccaforte W, Burke W, Bayer B, Wengel S. Validation of a Telephone Version of the Mini-Mental State Examination. *Journal of the American Geriatrics Society* 1992;40:697-702.
- (35) Anderson C, Laubscher S, Burns R. Validation of the Short Form 36 (SF-36) Health Survey Questionnaire Among Stroke Patients. *Stroke* 1996;27:1812-6.
- (36) Dorman P, Slattery J, Farrell B, Dennis M, Sandercock P. Qualitative Comparison of the Reliability of Health Status Assessments with the EuroQol and the SF-36 Questionnaires after Stroke. *Stroke* 1998;29:63-8.
- (37) Dorman P, Waddell F, Slattery J, Dennis M, Sandercock P. Is the EuroQol a Valid Measure of Health-Related Quality of Life After Stroke? *Stroke* 1997;28:1876-82.
- (38) Finch E, Brooks D, Stratford P, Mayo N. *Physical Rehabilitation Outcome Measures*. 2nd ed. Hamilton: BC Decker Inc; 2002.

- (39) Grootendorst P, Feeny D, Furlong W. Health Utilities Index Mark 3: Evidence of Construct Validity for Stroke and Arthritis in a Population Health Survey. *Medical Care* 2000;38(3):290-9.
- (40) Mathias S, Bates M, Pasta D, Cisternas M, Feeny D, Patrick D. Use of the Health Utilities Index with Stroke Patients and Their Caregivers. *Stroke* 1997;28(10):1888-94.
- (41) Duncan P, Wallace D, Lai S, Johnson D, Embretson S, Laster L. The Stroke Impact Scale Version 2.0. *Stroke* 1999;30:2131-40.
- (42) Duncan P, Lai S, Tyler D, Perera S, Reker D, Studenski S. Evaluation of Proxy Responses to the Stroke Impact Scale. *Stroke* 2002;33:2593-9.
- (43) Poissant L, Mayo N, Wood-Dauphinee S, Clarke A. The Development and Preliminary Validation of a Preference-Based Stroke Index (PBSI). *Health and Quality of Life Outcomes* 2003 September 10;1(43).
- (44) Collin C, Wade D, Davies S, Horne V. The Barthel Index:a Reliability Study. *International Disability Studies* 1988;10:61-3.
- (45) Wade D, Collin C. The Barthel Index: a Standard Measure of Physical Disability? *Disablity Studies* 1988;10:64-7.

Chapter 6: Linking the Statistical Framework for Examining Response Shift and the Identification of Response Shift in Physical Function

The previous chapter was a manuscript describing a framework to examine response shift statistically when greater than two time points are available for analysis. The example used in the framework was mental health, in which response shift was identified after stroke, primarily at the 12 month time-frame. The framework was valuable in guiding the search for the identification of the localization of response shift.

The next chapter uses this same methodology to estimate the extent to which response shift occurs in physical function; the manuscript in Chapter 7 is to be submitted to the clinical journal, 'Stroke' in September 2008. We originally hypothesized that response shift would not occur in the physical function construct but this hypothesis was not supported. Physical function was identified as a response shift susceptible construct of HRQL. As in the mental health construct, response shift was identified primarily at the 12 month period. The use of the word 'difficulty' in the response options occurred in two of the measured variables in which response shift was identified. It is possible that wording and framing of the response options may be susceptible to response shift.

Chapter 7: Can Response Shift Occur in Self-Perceived Physical Function? Using Stroke as a Model

To be submitted to 'Stroke'.

Ruth Barclay-Goddard BMR(PT), MHSc¹

Lisa Lix PhD²

Robert Tate PhD³

Leah Weinberg PhD¹

Nancy E. Mayo, BSc(PT), MSc, PhD⁴

1. Department of Physical Therapy, School of Medical Rehabilitation, University of Manitoba, R106-771 McDermot Ave. Winnipeg, Manitoba, Canada R3E 0T6 barclayg@cc.umanitoba.ca Phone (204)977-5637, fax (204)789-3927

2. School of Public Health, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

3. Department of Community Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada

4. Department of Medicine and School of Physical and Occupational Therapy, McGill University; Division of Clinical Epidemiology; Division of Geriatrics, McGill University Health Center, Royal Victoria Hospital Site, R4.29, 687 Pine Ave W, Montreal, QC, H3A 1A1 nancy.mayo@mcgill.ca Phone (514) 934 1934 36922, fax (514) 843 1493

Abstract

Background and Purpose: Response shift (RS) is a change of the self-perceived meaning of a construct resulting from change in internal standards, values, or conceptualization. Presence of RS suggests individuals use different metrics over time, invalidating the use of change score evaluations to assess health status. The objective of this study was to determine if RS was present in a model of physical function (PF) over time post stroke. **Methods:** Data from the multi-centre longitudinal study “Understanding Quality of Life Post-Stroke: A Study of Individuals and their Caregivers” was analyzed. 678 persons at 1, 3, 6, and 12 months post stroke participated. Mean age was 67 years, 45% of the participants were female. PF was assessed with subscales from the SF-36, Euroqol, Stroke Impact Scale, Preference Based Stroke Index, and the Health Utilities Index. Structural equation modeling was utilized to determine if RS occurred over time. **Results:** A chi-square difference test between constrained and unconstrained longitudinal models suggested RS. Reprioritization RS was noted in physical activities; recalibration RS was observed in subscales of physical activities, stairs, walking, and hand function, primarily at 12 months. **Conclusions:** The observation of RS was unexpected and has implications for change measurement in PF. Measures that focus on ‘difficulty’ in task performance may be sensitive to RS, as opposed to performance based measures. This has implications for choosing self-perceived or performance based measures to detect change in physical function over time.

Key Words:

Stroke

Physical function

Structural equation modeling

Response shift

Health related quality of life

Word Count:

Text = 3462

Figure 1 = 38

ref = 1089

abstract = 236

key words = 15

title page = 146

total = 4986

Presented in part as a platform presentation at the Canadian Physiotherapy Association Congress in Ottawa, Canada in 2008.

The data source from the original longitudinal study, “Understanding Quality of Life Post-Stroke: A Study of Individuals and their Caregivers” was funded by the Canadian Stroke Network.

Introduction

Health related quality of life (HRQL) is self perceived and can change over time. It is defined as “the value assigned to duration of life as modified by the impairments, functional states, perceptions and social opportunities influenced by disease, injury, treatment or policy”¹. Increased HRQL is the ultimate goal in the rehabilitation of individuals with a variety of health conditions, including stroke². There are three domains that are commonly used to describe and estimate HRQL: physical, psychological, and social functioning³.

One of challenges with evaluating HRQL over time is that people can recalibrate their ratings based on personal and vicarious experience, they can change their priorities for different aspects of HRQL and the whole construct can even be reconceptualized. These types of changes are not true change and have been described as response shift.

This paper is part of a larger study assessing response shift in construct specific HRQL models over time post stroke. These models include physical function, mental health, and participation. This paper describes the model of physical function. It was hypothesized that parameter estimates in the physical function model would not change over time; response shift would not be detected. Aspects of physical function are typically not expected to undergo response shift⁴.

Response shift is evident in individuals with chronic illnesses or disability. Many will report HRQL similar to that of healthy individuals, despite ongoing activity limitations^{2,5,6}. Response shift is defined as “a change in the meaning of one’s self-evaluation of a target construct as a result of: (a) a change in the respondent’s internal standards of

measurement (i.e. scale recalibration); (b) a change in the respondent's values (i.e. the importance of component domains constituting the target construct) or (c) a redefinition of the target construct (i.e. reconceptualization)"⁷.

In recalibration, the individual changes his / her internal measurement standards, comparable to a ruler that shrinks or stretches ⁸. For example, a young woman fractures her tibia and describes the resulting pain as a 10/10. Two years later, when in labour, she realizes that the leg pain was really only a 9/10 and the labour pain was a 10/10. Non-uniform recalibration occurs when a portion of the measurement scale is recalibrated; uniform recalibration occurs when the entire scale is recalibrated.

Reprioritization is demonstrated when a change occurs in the order of the importance of domains making up HRQL ⁷. This is demonstrated in an individual for whom important aspects of HRQL include sports participation, cognitive function, and family, in that order. After a mild stroke, the importance of domains changes to: family, cognitive function, and sports participation.

Reconceptualization involves a redefinition of HRQL ⁷. After rehabilitation for a spinal cord injury, an individual may consider the ability to work as very important to his HRQL. Prior to his injury, however, he had not even considered that the ability to work affected his HRQL. The domains that describe HRQL for this individual changed, reflecting reconceptualization. With all types of response shift, subsequent testing reveals that the understanding of the construct being tested changes for the individual.

Response shift is likely to occur if change in health is recent, intense and all-encompassing ⁹. It has been suggested that recalibration is more likely in the first few

months after a threatening event and that clients with more severe symptoms engage in recalibration for longer time intervals than those with milder symptoms¹⁰. Stroke is a health condition that is ideal for studying response shift because it may produce sudden and intense changes in function, perception, cognition, mood, speech and HRQL⁵.

Initial studies in the evaluation of response shift post stroke have both suggested and refuted the presence of response shift. Three methods of response shift identification have been studied. The 'then test' method suggested recalibration in the EuroQol visual analogue scale and the Patient Generated Index (PGI) method suggested reconceptualization and reprioritization^{11, 12}. Structural equation modeling (SEM), using a measurement model based on the Short Form 36 (SF-36), did not demonstrate response shift¹³. SEM has been used in other client populations to determine the presence of response shift¹⁴⁻¹⁶. Response shift may not have been detected previously with SEM post stroke due to the time-frame (1 to 6 months), the measurement model developed, or the SEM methodology used.

SEM is the statistical approach used to determine response shift in this study. It is beneficial in situations when using secondary data analyses and when incorporating additional measures into a study to detect response shift is not possible. The particular SEM approach used in this study is suitable because it detects all 3 types response shift, unlike a previously described SEM approach^{17, 18}. Other statistical approaches are available, but are unable to identify reconceptualization, as well as reprioritization, and recalibration.

In SEM, latent variables are associated with measured variables which reflect the construct of interest. A latent variable is a construct which the researcher is interested in measuring, but can not truly be measured directly, such as HRQL. Descriptions of SEM can be found elsewhere ¹⁹.

We hypothesized that there would be no response shift in physical function post stroke. Aspects of HRQL which are objective and directly measurable (e.g. physical function tasks such as the ability to walk 20 metres independently) are thought to be unlikely to undergo response shift ⁴.

Methods and Data analysis

This study involves secondary analysis from the observational study “Understanding Quality of Life Post-Stroke: A Study of Individuals and their Caregivers” ²⁰. This multi-centre project accumulated data from 678 persons at 1, 3, 6, and 12 months post stroke. Three hundred and forty-two participants (50.4%) had complete data at all four measurement occasions, and 454 participants (67.0%) completed the study at 12 months. Generic HRQL indices in the study were the Short-Form 36 (SF-36) ²¹, the EuroQuol (EQ5D) ²², and the Health Utilities Index (HUI) ²³. Stroke-specific HRQL indices included the Stroke Impact Scale (SIS) ²⁴ and the Preference-Based Stroke Index (PBSI) ²⁵. All of these indices have been validated with individuals with stroke ²¹⁻²⁵. Ethics approval for secondary analysis was obtained from the Health Research Ethics Board at the University of Manitoba.

Study participants were characterized with cross-sectional descriptive statistics at four time points with missing data excluded, to describe personal factors, disability related

factors, and HRQL indices. Variables were characterized with tests for normality of the subscales of HRQL indices. Missing data was evaluated using χ^2 comparison of specific outcomes and missingness at any time to determine the missing data pattern. Data screening and data preparation was completed with SAS 9.14²⁶; SEM analysis utilized LISREL 8.72, SIMPLIS²⁷.

The SEM method described by Oort, was used to identify the presence of response shift in the longitudinal data¹⁷. Oort described four steps to follow when using SEM to evaluate response shift¹⁷. The first step is to determine an appropriate measurement model. The second step is an overall test of response shift. To determine if response shift is present in the longitudinal model, a model with all parameters freely estimated is compared to a constrained model in which the error variances, paths, and intercepts are constrained to be equal across time. The chi-square (χ^2) difference test evaluates the change in overall fit of the free and fully constrained (no response shift) models¹⁹. When the χ^2 difference test is statistically significant, the null hypothesis that the two models fit the same is rejected. If they fit differently, response shift is present in the model. The third step is the detection of the type and location of response shift. Oort describes an assessment of true change (change of the latent variables over time) in the last step¹⁷. In step three, a change in the pattern of factor loadings suggests reconceptualization. For example, a measured variable has an association (path/factor loading) with a latent variable at time 2 but not at time 1. A change in the magnitude of factor loadings implies reprioritization (e.g. the strength of association between the measured variables and the latent variables over time). An intercept change over time suggests uniform recalibration; non-uniform recalibration is implied by a change in error variances¹⁷.

A framework has also been developed by the authors to aid in guiding the search for the localization of response shift with SEM with greater than two time points²⁸. This framework considers important issues such as the purpose of the study, validation of the SEM model, and controlling Type I error (determining that there is a difference when there is not). The purpose may be exploratory to identify localized sources of response shift to inform future research or theory-driven to confirm the localized sources of response shift from previous studies²⁸. Validation of a model increases confidence in the model and the resulting identification of response shift. In this study, the longitudinal model was validated by comparing full information maximum likelihood (FIML) estimation, used with missing data, and robust maximum likelihood estimation (MLE), used for non-normal data. Control of familywise Type I error was also considered. When multiple inferences (a 'family' of inferences) are made in a single study, a common method to control the type1 error rate is to adjust p values with the Bonferroni correction²⁹. The α significance level (0.05) was divided by the number of models tested.

A measurement model that reflected physical function at one month was developed; this cross sectional model was utilized to form the four-time longitudinal model. FIML was the parameter estimation method used. To determine the content of the physical function model, we used a description of physical health from Stewart and Ware: self care / ADL, moving and walking (including upper and lower extremity function and transfers), pain , energy / fatigue , and sleep³⁰. Pain is not thought to be a large part of post stroke recovery, except for those individuals with post-stroke shoulder pain. In a group of community dwellers at 6 months post stroke, pain scale values on the SF-36 were deemed to be equivalent to controls². Pain was subsequently not included in the model. There

were no variables related to sleep, and only one related to energy, therefore no associated latent constructs were developed. Items and subscales reflecting the latent constructs of self care, and mobility (moving and walking) were available in the HRQL indices in the study. Measured variables that reflect the self care and mobility latent variables are described in Table 7-1.

As has been suggested, numerous fit indices were used to describe the overall fit of the model to the sample ¹⁹. The null hypothesis of the χ^2 test is that the model has a perfect fit; a non-significant χ^2 suggests a good model fit ¹⁹. Models with large sample sizes, however, can often be rejected because the model χ^2 is affected by sample size ³¹. For this reason, some researchers use a normed χ^2 (χ^2/df) as a way to decrease the effect of the sample size on the χ^2 ; values up to 5.0 suggest a reasonable model fit ¹⁹. The root mean square error of approximation (RMSEA) of ≤ 0.05 suggests a close fit, between 0.05 and 0.08 is considered an acceptable fit, while $RMSEA \geq 1.0$ indicates a poor fit ³². LISREL also provides an estimate of fit improvement (modification indices) which would occur if specific parameters in the measurement model were changed.

Results

One of the 678 participants admitted into the study had no data collected at the four time points; therefore, data from 677 participants was analyzed. Demographics of the study participants are described in Table 7-2. Statistical data screening using χ^2 comparison of specific outcomes suggested the missing data pattern of missing at random, based on co-variates. All variables used were moderately non-normal. The characteristics of the study participants on the measured HRQL variables used in the model at all four time points are detailed in Table 7-3. The values presented in table 7.3 are calculated from observed data

only with no imputation. The covariance and means matrices produced by FIML are available on request from the authors.

A measurement model was constructed with the selected variables. Due to suggestions from the modification indices, the dexterity, hand, and strength variables were associated with the self-care latent variable instead of the mobility latent variable. It is sensible that good hand function / dexterity and strength of the limbs would assist with ADL tasks. The SIS ADL subscale was also associated with the mobility latent variable; the SIS ADL subscale includes instrumental ADL items which require a fair bit of mobility. Error correlations were added between variables that measured similar constructs. The cross sectional model for one month is presented in Figure 7-1. The model has an acceptable fit with RMSEA of 0.076 (90% confidence interval: 0.067 to 0.085) but χ^2 fit statistics provided conflicting appraisal of the fit as is common with large samples (χ^2 270.00, 59 degrees of freedom (df), $p=0.00$). Normed $\chi^2= 4.58$, which can be considered reasonable. Model fit in this study focused on the RMSEA and normed χ^2 due to the common poor fit of χ^2 in large samples. Only χ^2 and RMSEA fit statistics were available with FIML in LISREL. A four time point longitudinal model was developed from the one month model. The longitudinal model demonstrated a close fit with an RMSEA of 0.046 (0.044 to 0.048); χ^2 2800.31, 1148 df, $p=0.00$. Normed $\chi^2= 2.44$, which is reasonable.

Model validity was confirmed by comparing FIML and a robust MLE. Both methods resulted in reasonable to close fitting longitudinal models. [See Appendix 7-1, included in thesis but not in paper for clinical journal].

The χ^2 difference test between the free and fully constrained longitudinal models was 507.90 with 108 df, significant at $p < 0.001$. This suggests that, in contrast to our hypothesis, response shift occurred somewhere in the model. Based on this finding, we developed a post hoc hypothesis that response shift would be identified in measured variables where descriptors of ‘difficulty’ were present. The concept of ‘difficulty’ could potentially undergo changes in interpretation over time.

All possible sources of response shift were tested with an exploratory approach to identify the localization of response shift. Modification indices suggested removals of constraints, eight of which were sensible based on knowledge of stroke and response shift in general (i.e., not removing constraints at baseline). Constraints were removed from the constrained model one by one to identify where response shift was located ¹⁷.

Response shift was identified in 5 locations (4 items / subscales): reprioritization of physical activities around 12 months, uniform recalibration of physical activities and stair walking around 12 months, uniform recalibration of hand function around 3 months, and non-uniform recalibration of walking around 12 months. The word ‘around’ is used, as it is not known exactly when response shift occurred, but it may have occurred before or after (around) the period tested. Table 7-4 demonstrates changes in the parameter estimates of these variables.

The response shift around 12 months occurred in items of the PBSI, and 3 month response shift occurred in the SIS. Two of the four items that demonstrated response shift used the term ‘difficulty’ in the descriptors: PBSI walking and SIS hand function. The

effect size (ES) of observed (measured) change, response shift, and true change were calculated as described by Oort¹⁷ and are presented in Table 7-5.

Discussion

Of the five locations of response shift identified, four of those occurred around 12 months. It would appear that if response shift is to occur in physical function post stroke, it occurs later. A previous study did not identify response shift in physical and mental health. Perhaps the timeframe of one to six months was not long enough¹³.

The ES of true change in response shift susceptible variables was small, while the ES of response shift was negligible (< 0.20). However, observed change in physical activity had a moderate ES which became small when response shift was removed. Future studies about the clinical importance of response shift are required.

Four of the five locations where response shift was identified involved three items from the PBSI. One of the items, physical activities, demonstrated 2 types of response shift: reprioritization and uniform recalibration. The PBSI is the first preference based measure specifically developed for stroke²⁵.

We had not expected to identify response shift in physical function. However, we then hypothesized that the wording of some measured variables might be amenable to response shift, such as the term 'difficulty' in a number of response options. Of the four items where response shift was identified, two had the term 'difficulty' in the response options (SIS hand function and PBSI walking). There is a difference in asking whether a person does an activity or has difficulty with the activity; some individuals continue to be physically active despite difficulties while others are inactive despite few difficulties³³.

Schwartz and Rapkin described the difference between performance based measures (such as a timed walking test), perception based measures (such as an individual describing how many times a day they walk), and evaluation based / self perceived measures (i.e. how difficult is it to walk?)³⁴. Changes over time that are not true change may be: measurement error in performance based measures, response bias in perception based measures, or response shift in evaluation-based methods³⁴. The evaluation of ‘difficulty’ is self perceived and therefore susceptible to response shift. Functional health has been defined as “the ability of an individual to perform and adapt to one’s environment, measured both objectively and subjectively over a given period”³⁵. Performance based and self-reported measures look at diverse aspects of function and both should be included when measuring function³⁶. Future studies of response shift in physical function should evaluate models that incorporate both objective (performance based) and subjective (self-perceived) measures. Alternately, a comparison of models of physical function with only performance based measures and models with self perceived measures would be of interest in stroke and other chronic conditions.

Pain was not included in this model. Studies on stroke complications have suggested the frequency of shoulder pain is 2% and pain in other areas is 24% during the first week post stroke³⁷; 15% and 41% respectively from discharge to six months³⁸. The issue of response shift in pain post stroke deserves further evaluation.

There are limitations present. SEM is a large sample method. There were data from 677 individuals in this study; it is suggested that the number of cases to free parameters estimated should be at least 10:1¹⁹. With the large longitudinal model, there are over 100 parameters estimated; the sample size may appear too small. To deal with potential

overfitting of the data (many parameters with less than ideal number of participants), we validated the results using more than one estimation method; the results appeared to be robust. Since SEM involves group level analysis, results are averaged across all individuals. Not all individuals would be expected to undergo response shift and those who do experience response shift may not all experience it at the same time, which may influence results.

Conclusion

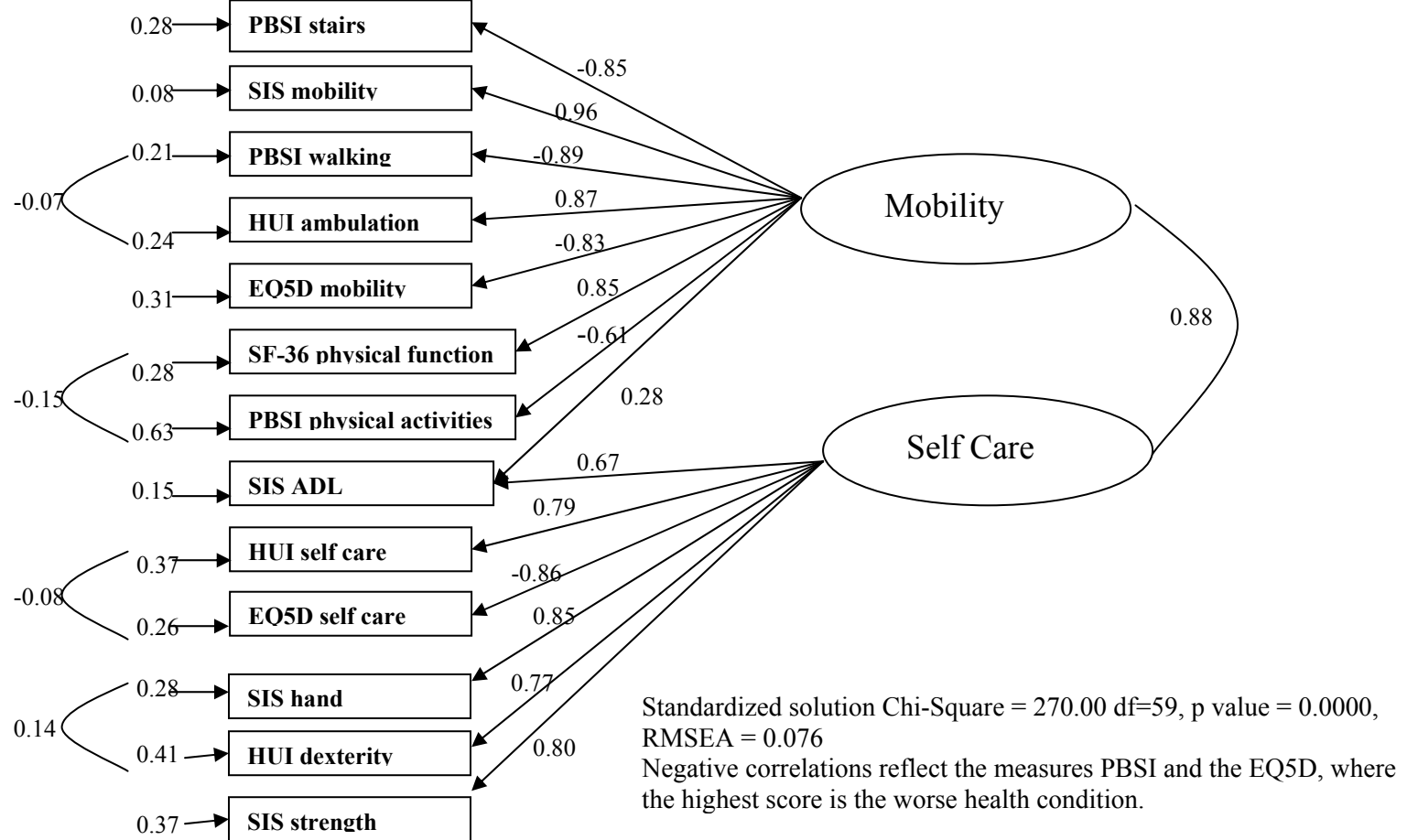
We believe that this is the first description of a longitudinal model of physical function post stroke as well as the first identification of response shift in physical function over one year post stroke. This finding was unexpected, but may be partly explained by the evaluation of the concept of ‘difficulty’ in some of the items. The findings in this study can serve to inform future research of response shift in physical function post stroke.

Response shift presence in physical function will affect the way in which we evaluate physical function in the clinic and in research studies, the type of measures that we use, and the way in which change in physical function is analyzed. Careful consideration of the interpretation of the term difficulty would need to be made in self-report physical function indices that comprise items with the concept of difficulty as response options.

Response shift, as well as improved HRQL, is also a rehabilitation goal, though not typically stated in those terms by clinicians^{4, 6, 39, 40}. Facilitating response shift can be particularly important when impairments, activity limitations, and participation restrictions are not expected to recover fully, but improved HRQL is a goal. If measures of physical function demonstrate response shift, choice of self perceived or performance

based measures need to be carefully chosen by the clinician to avoid bias in change estimates. If response shift is a focus of treatment, self perceived measures of response shift may be most appropriate, as these measures appear susceptible to response shift. If the focus of evaluation is to measure true change in physical function, performance based measures may be most appropriate.

Figure 7-1 Physical Function Model – 1 month



Ovals represent latent variables of self care and mobility; rectangles represent measured variables from HRQL indices that are associated with the latent variables. The arrows between the latent and measured variables (factor loadings or paths) are as regression coefficients. Curved lines also represent correlations and the small arrows represent error variances.
 MMSE = Mini-Mental Status Exam, SIS = Stroke Impact Scale, EQ5D = Euroqol, SF-36 = Short Form 36, PBSI = Preference Based Stroke Index, HUI = Health Utilities Index

Table 7-1 Items and Subscales from HRQL Indices and Associated Latent Variables

Latent Variable	HRQL Indices				
	SF-36	EQ5D	HUI	SIS	PBSI
Self care		self care	self care	self care	
Mobility (moving and walking)	Physical function	mobility	ambulation, dexterity*	mobility, hand,*+ strength*	walking+, stairs, physical activity
Scoring	subscales 0-100 higher = better	items 1- 3 lower score = better	subscales 0-1 higher = better	subscales 0-100 higher = better	items 1- 3 lower score = better

SF-36 = Short Form 36, EQ5D = EuroQol, HUI = Health Utilities Index, SIS = Stroke Impact Scale, PBSI = Preference Based Stroke Index
 * items associated with self care in final model , + items using “difficulty” in the response options

Table 7-2 Demographics of Study Participants

Variable	N	Mean (SD)	Percent
Female	677		45
Age at stroke	676	68 (14.8)	
First stroke	657		94
Side of lesion	654		
Left			45
Right			50
Bilateral			5
Location of hemiplegia	464		
Lower Extremity			5
Upper Extremity			12
Upper and Lower Extremities			83
Barthel Index/100 - 1 month ⁴¹	668	75(25.9)	
Telephone Mini-Mental State Exam – 1 month - rescored/100 ⁴²	502	89(14.8)	

Table 7-3 Outcomes of Measured Variables over One Year Post Stroke

Variable	1 month			3 months			6 months			12 months		
	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD
EuroQol (EQ5D)												
Mobility	530	1.7	0.7	463	1.5	0.5	420	1.5	0.6	413	1.2	0.5
self care	529	1.5	0.6	461	1.3	0.5	420	1.3	0.5	413	1.2	0.5
Health Utilities Index 2 (HUI)												
self care	504	0.64	0.47	469	0.77	0.41	439	0.80	0.39	436	0.82	0.37
Health Utilities Index 3 (HUI)												
Ambulation	529	0.63	0.40	492	0.76	0.32	460	0.79	0.29	436	0.80	0.30
Dexterity	548	0.75	0.34	502	0.85	0.27	465	0.84	0.28	438	0.85	0.28
Preference Based Stroke Index (PBSI)												
Walking	596	1.8	0.8	529	1.5	0.7	478	1.5	0.6	450	1.4	0.6
Stairs	595	1.9	0.8	526	1.6	0.7	475	1.5	0.6	449	1.6	0.7
physical activities	596	2.6	0.6	528	2.4	0.7	475	2.4	0.7	446	2.3	0.7
Short Form 36 (SF-36)												
physical function	526	44.3	34.3	464	52.8	33.8	412	59.5	32.0	406	59.2	32.0
Stroke Impact Scale (SIS)												
ADL	611	66.7	27.1	529	77.1	22.9	480	78.9	22.0	445	81.3	21.5
hand	603	59.6	39.1	527	70.0	34.7	476	74.3	32.6	443	76.1	31.9
mobility	612	64.0	31.3	530	75.2	25.0	478	77.8	23.7	444	79.9	22.2
strength	606	64.8	29.0	527	70.7	26.1	478	71.6	25.9	442	73.9	25.4

Note: EQ-5D and PBSI – lower score is higher functioning; HUI, SIS, and SF-36 – higher score is higher functioning

Table 7-4 Localization of Response Shift - Changes in Non-Standardized Parameter Estimates

Constraint Removed	1 month	3 months	6 months	12 months	1df χ^2 difference	p value*
<i>Reprioritization</i>	Path loading					
PBSI physical activities	-0.01	-0.01	-0.01	-0.02	47.42	p<0.0001
<i>Uniform Recalibration</i>	Intercepts					
PBSI physical activities	2.63	2.63	2.63	2.54	16.66	p<0.0001
PBSI stairs	1.88	1.88	1.88	1.94	7.98	p=0.0047
SIS hand function	59.56	57.51	59.56	59.56	7.55	p=0.0060
<i>Non-uniform Recalibration</i>	Error variances					
PBSI walking	0.13	0.13	0.13	0.01	9.48	p=0.0021

* α significance level $p \leq 0.0063$ Response shift is demonstrated in those indices with a parameter estimate that is different at a particular time-frame (bolded). SIS = Stroke Impact Scale, PBSI = Preference Based Stroke Index

Table 7-5 Effect Sizes of Observed Change, Response Shift, and True Change

Variable	Effect Sizes*				
	Type of response shift	Comparison	Observed changes	Response shift contribution	True change contribution
Preference Based Stroke Index					
Walking	non uniform recalibration	1 & 12 months	-0.41	**0.00	-0.44
Stairs	uniform recalibration	1 & 12 months	-0.38	0.06	-0.44
physical activity	uniform recalibration	1 & 12 months	-0.50	-0.14	-0.35
	reprioritization			-0.15	
Stroke Impact Scale					
Hand	uniform recalibration	1 & 3 months	0.23	-0.05	0.28
*ES has been described as small (0.2), medium (0.5) and large (0.8) ⁴³ .					
**non-uniform recalibration was identified, but non-uniform recalibration is not estimated in the effect size calculations					

Appendix 7-1 Physical Function Model Validity (sensitivity analysis) for Longitudinal Model

Method	χ^2	df	Normed χ^2	RMSEA (90%CI)	CFI	S-B χ^2	Normed S-B χ^2	SRMR
FIML	2800.29	1148	2.44	0.046 (0.044 ; 0.048)	---	---	---	---
Imputed data sets with robust maximum likelihood estimation								
#1	5479.06	1148	4.77	0.060 (0.058 ; 0.062)	0.99	3981.81	3.47	0.053
#2	5224.18	1148	4.55	0.058 (0.056 ; 0.060)	0.99	3750.11	3.27	0.052
#3	5466.74	1148	4.76	0.061 (0.059 ; 0.063)	0.99	4063.90	3.54	0.061
#4	5525.76	1148	4.81	0.060 (0.058 ; 0.062)	0.99	3983.29	3.47	0.052
Mean (SD)	5423.94 (135.57)	1148	4.72	0.060 (0.001)	0.99	3944.78 (135.32)	3.44	0.055 (0.004)

Normed χ^2 (χ^2/df) - values up to 5 indicate reasonable fit
 RMSEA \leq 0.05 close fit; 0.05 to 0.08 reasonable fit; \geq 0.10 poor fit
 CFI $>$ 0.90 reasonably good fit, SRMR $<$ 0.10 is favourable¹⁹

References

- (1) Patrick D, Erickson P. *Health Status and Health Policy - Quality of Life in Health Care Evaluation and Resource Allocation*. Oxford: Oxford University Press; 1993.
- (2) Mayo N, Wood-Dauphinee S, Cote R, Durcan L, Carlton J. Activity, Participation, and Quality of Life 6 Months Poststroke. *Archives of Physical Medicine and Rehabilitation* 2002;83:1035-42.
- (3) Sprangers MA. Quality-of-life assessment in oncology. Achievements and challenges. *Acta oncologica Stockholm, Sweden* 2002;41(3):229-37.
- (4) Wilson IB. Clinical Understanding and Clinical Implications of Response Shift. *Social Science and Medicine* 1999;48:1577-88.
- (5) Mayo N, Wood-Dauphinee S, Ahmed S, Gordon C, Higgins J, McEwen S, Salbach N. Disablement Following Stroke. *Disability and Rehabilitation* 1999;21(5/6):258-68.
- (6) Schwartz C, Andresen EM, Nosek M, Krahn G, RRTC Expert Panel on Health Status Measurement. Response Shift Theory: Important Implications for Measuring Quality of Life in People With Disability. *Archives of Physical Medicine and Rehabilitation* 2007;88:529-36.

- (7) Schwartz CE, Sprangers MA. Methodological Approaches for Assessing Response Shift in Longitudinal Health-Related Quality of Life Research. *Social Science and Medicine* 1999;48:1531-48.
- (8) Armenakis A. A Review of Research on the Change Typology. In: Pasmore W, Woodman R, editors. *Research in Organizational Change and Development*. 2 ed. Greenwich: JAI Press Inc.; 1988. p. 163-94.
- (9) Sprangers MA, Schwartz CE. Integrating Response Shift into Health-Related Quality of Life Research: A Theoretical Model. *Social Science and Medicine* 1999;48:1507-15.
- (10) Schwartz CE, Sprangers MAG, Carey A, Reed G. Exploring response shift in longitudinal data. *Psychology and Health* 2004 February;Vol 19(1):51-69.
- (11) Ahmed S, Mayo N, Wood-Dauphinee S, Hanley J, Cohen S. Response Shift Influenced Estimates of Change in Health-Related Quality of Life Poststroke. *Journal of clinical epidemiology* 2004;57:561-70.
- (12) Ahmed S, Mayo N, Corbiere M, Wood-Dauphinee S, Hanley J, Cohen R. Using the Patient Generated Index to Evaluate Response Shift Post Stroke. *Quality of Life Research* 2005;14:2247-57.
- (13) Ahmed S, Mayo N, Corbiere M, Wood-Dauphinee S, Hanley J, Cohen R. Change in Quality of Life of People with Stroke Over Time: True Change or Response Shift? *Quality of Life Research* 2005;14:611-27.

- (14) Oort FJ, Visser M, Sprangers MA. An Application of Structural Equation Modeling to Detect Response Shifts and True Change in Quality of Life Data from Cancer Patients Undergoing Invasive Surgery. *Quality of Life Research* 2005;14:599-609.
- (15) Visser M, Oort F, Sprangers M. Methods to Detect Response Shift in Quality of Life Data:A Convergent Validity Study. *Quality of Life Research* 2005;14:629-39.
- (16) Nolte S, Elsworth G, Sinclair A, Osborne R. Replication of a Structural Equation Modeling Approach to Detect Response Shift in Participants of Health Education Programs. *Quality of Life Research* October[suppl.], A-13. 2007.
Ref Type: Abstract
- (17) Oort FJ. Using Structural Equation Modeling to Detect Response Shifts and True Change. *Quality of Life Research* 2005;14:587-98.
- (18) Schmitt N. The Use of Analysis of Covariance Structures to Assess Beta and Gamma Change. *Multivariate Behavioural Research* 1982;17:343-58.
- (19) Kline R. *Principles and Practice of Structural Equation Modeling*. 2 ed. New York: The Guilford Press; 2005.
- (20) Wood-Dauphinee S. Quality of Life Post Stroke. *Brainwaves - newsletter for members of the Canadian Stroke Network* 2003;Issue 19(May):2-4.

- (21) Anderson C, Laubscher S, Burns R. Validation of the Short Form 36 (SF-36) Health Survey Questionnaire Among Stroke Patients. *Stroke* 1996;27:1812-6.
- (22) Dorman P, Waddell F, Slattery J, Dennis M, Sandercock P. Is the EuroQol a Valid Measure of Health-Related Quality of Life After Stroke? *Stroke* 1997;28:1876-82.
- (23) Mathias S, Bates M, Pasta D, Cisternas M, Feeny D, Patrick D. Use of the Health Utilities Index with Stroke Patients and Their Caregivers. *Stroke* 1997;28(10):1888-94.
- (24) Duncan P, Wallace D, Lai S, Johnson D, Embretson S, Laster L. The Stroke Impact Scale Version 2.0. *Stroke* 1999;30:2131-40.
- (25) Poissant L, Mayo N, Wood-Dauphinee S, Clarke A. The Development and Preliminary Validation of a Preference-Based Stroke Index (PBSI). *Health and Quality of Life Outcomes* 2003 September 10;1(43).
- (26) SAS Institute. *SAS/STAT User's Guide, Version 9*. Cary,NC: SAS Institute, Inc.; 2004.
- (27) Jöreskog K, Sörbom D. *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Lincolnwood, IL: Scientific Software International Inc.; 1993.
- (28) Barclay-Goddard R, Lix L, Mayo N, Wood-Dauphinee S, Tate R. A Framework for Assessing Response Shift in Post -Stroke Health Related Quality of Life Over

- Multitple Time Points. Quality of Life Research October[suppl.], A-14. 2007.
- Ref Type: Abstract
- (29) Ludbrook J. Statistics in Physiology and Pharmacology: A Slow and Erratic Learning Curve. *Clinical and Experimental Pharmacology and Physiology* 2001;28:488-92.
- (30) Stewart A. The Medical Outcomes Study Framework of Health Indicators. In: Stewart A, Ware J, editors. *Measuring Functioning and Well-Being*. Durham: Duke University Press; 1992. p. 12-24.
- (31) Hays R, Revicki D, Coyne K. Application of Structural Equation Modeling to Health Outcomes Research. *Evaluation and the Health Professions* 2005;28(3):295-309.
- (32) Browne M, Cudeck R. Alternative Ways of Assessing Model Fit. In: Bollen K, Long J, editors. *Testing Structural Equation Models*. London: Sage Publications; 1993. p. 136-62.
- (33) Rantanen T, Guralnik J, Sakari-Rantala R, Leveille S, Simonsick E, Ling S, Fried L. Disability, Physical Activity, and Muscle Strength in Older Women: The Women's Health and Aging Study. *Archives of Physical Medicine and Rehabilitation* 1999;80:130-5.
- (34) Schwartz CE, Rapkin BD. Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. *Health and quality of life outcomes [electronic resource]* 2004 March 23;2(1):16.

- (35) Sousa K, Kwok O. Putting Wilson and Cleary to the Test: Analysis of a HRQOL Conceptual Model Using Structural Equation Modeling. *Quality of Life Research* 2006;15:725-37.
- (36) Reid A, Birmingham T, Stratford P, Alcock G, Giffin J. Hop Testing Provides a Reliable and Valid Outcome Measure During Rehabilitation After Anterior Cruciate Ligament Reconstruction. *Physical Therapy* 2007;87(3):337-49.
- (37) Indredavik B, Rohweder G, Naalsund E, Lydersen S. Medical Complications in a Comprehensive Stroke Unit and an Early Supported Discharge Service. *Stroke* 2008;39:414-20.
- (38) Langhorne P, Stott D, Robertson L, MacDonald J, Jones L, McAlpine C, Dick F, Taylor G, Murray G. Medical Complications After Stroke - A Multicenter Study. *Stroke* 2000;31:1223-9.
- (39) Osborne R, Hawkins M, Sprangers M. Change of Perspective: A Measureable and Desired Outcome of Chronic Disease Self-Management Intervention Programs That Violates the Premise of Preintervention / Postintervention Assessment. *Arthritis and Rheumatism* 2006;55(3):458-65.
- (40) Rochette A, Korner-Bitensky N, Levasseur M. 'Optimal' Participation: A Reflective Look. *Disability and Rehabilitation* 2006;28(19):1231-5.
- (41) Wade D, Collin C. The Barthel Index: a Standard Measure of Physical Disability? *Disability Studies* 1988;10:64-7.

- (42) Roccaforte W, Burke W, Bayer B, Wengel S. Validation of a Telephone Version of the Mini-Mental State Examination. *Journal of the American Geriatrics Society* 1992;40:697-702.
- (43) Cohen J. *Statistical Power Analysis for the Behavioural Sciences*. 2nd ed. Hillsdale NJ: Erlbaum; 1988.

Chapter 8: The Participation Model

Link to Chapter 8

Chapter 1 discussed that the three constructs commonly used to estimate HRQL include physical function, mental health, and social functioning (participation) and that the impression of overall health / general well being should also be included in HRQL assessment¹. The identification of response shift in the three HRQL constructs of mental health, physical function, and participation is the focus of this thesis.

Chapter 5 described a framework that was developed to assist in identifying response shift over multiple time points with SEM. It was demonstrated that uniform and non-uniform recalibration response shifts were identified in the mental health construct of HRQL in six locations. Five localizations of response shift were identified at the 12 month period, suggesting response shift occurring between the six and twelve month time.

Chapter 7 described the identification of response shift in the physical function construct of HRQL, which had not been expected. A perception or evaluation of difficulty in completing an item may have partly explained the response shift that was identified. Again, a majority of the response shift identified in physical function was noted at the 12 month time period.

This chapter describes the third construct of HRQL – participation. The construct of overall health / general well-being was incorporated into a model of participation. Response shift was suggested in participation over one year post stroke. The identification of the localization of response shift in participation will be a focus of future

research. Since this chapter is not written as a manuscript, reference is made to information in previous chapters, specifically Chapters 1 and 3.

Participation

Participation restrictions are common post stroke²⁻⁴. The most commonly occurring participation restrictions and activity limitations in community dwelling individuals post stroke include: having a meaningful activity during the day, doing household tasks, traveling, and activities of daily living⁵. Participation has not been a major focus of research post stroke, but this is starting to change³.

In this thesis, models of the health related quality of life (HRQL) constructs of mental health and physical function have demonstrated response shift (see Chapters five and seven). The third construct commonly used to estimate HRQL is social functioning¹. For the purpose of this study, social functioning has been redefined as the broader construct of participation. As part of the International Classification of Functioning (ICF), participation is defined as ‘involvement in a life situation’⁶. Participation can also be defined as “the accomplishment and engagement of a person in his/her daily activities and social roles, resulting from the interaction between personal factors and environmental factors acting as facilitators or obstacles”³. Participation includes actions and roles that an individual requires for one’s well-being; these roles and activities vary from person to person depending on what is of value to them⁷.

There is consensus that the impression of overall health / general well being should also be included in HRQL assessment¹. In a study evaluating the association between instrumental activities of daily living (IADL) and self perceived well-being at 6 months

post stroke, leisure activities (an aspect of participation) were associated with well-being (satisfaction) ⁴. In that study, leisure activities included reading, using the telephone, writing letters, going out for social activities, gardening, and driving a car ⁴. Structural equation modeling (SEM) was utilized, however, the sample size (64) was small for an SEM analysis.

In a study using SEM with 591 older adults with stroke living in the community, a positive effect of activities of daily living (ADL) and IADL on self rated health was demonstrated⁸. Self rated health included a general health rating and an assessment of change in health ⁸. Other constructs affecting self rated health included the presence of chronic conditions and depressive symptomatology ⁸.

Rochette et al in 2006 proposed that response shift likely occurs in participation after stroke; expectations about participation are redefined by an individual, likely leading to response shift ⁷. The objective of this study is to determine if response shift is present in the HRQL construct participation over one year post stroke.

Methods

This study involved secondary data analysis of the cohort study “Understanding Quality of Life Post-Stroke: A Study of Individuals and their Caregivers”⁹; the data source is described in detail in Chapter 3. Data screening and data preparation was completed with SAS 9.14 ¹⁰; SEM analysis utilized LISREL 8.72, SIMPLIS and PRELIS ¹¹.

The SEM method used to identify response shift, as described by Oort was utilized ¹².

This method is described in Chapter 3. A measurement model that reflected participation

at time one (one month) was developed; this model was used to form a longitudinal model including all four measurement occasions (one, three, six, and 12 months).

Content of the latent variables in the participation model were based in part on the Medical Outcomes Study Framework of Health Indicators description of social functioning: social function, role function, and role limitations due to emotional problems or to physical health¹³. Due to the research that suggests a relationship between participation and self-rated health/well-being, a latent variable of self-perceived health and recovery was also included. This latent variable was named ‘health efficacy’. Table 8-1 lists items from each HRQL index in the study used to build the participation model.

Table 8-1 Items and subscales from HRQL indices and related constructs

Measure	Participation			
	Social function	Role Function	Restricted Roles	Health Efficacy
Short Form 36	Social function	-	Role physical Role emotional	General health
EuroQol	-	Usual activities		VAS (visual analogue scale)
Health Utilities Index	-	-	-	-
Stroke Impact Scale	Social participation	-	-	VAS- Global recovery
Preference Based Stroke Index	Recreational activities	Driving Work	-	-

Results

Data from 677 participants were analyzed. See Supplemental Table 3-2 in Chapter 3 for a description of the study participants. At one month post stroke, 45% of the participants were living at home, 49% were in rehabilitation, and less than 1% were in long term care

facilities. Data screening suggested the missing data pattern of missing at random; all variables used were moderately non-normal. See Chapter 3 for description of missing data and normality of data. The characteristics of the study subjects on the measured variables used in the model at all four time points are shown in Table 8-2.

Table 8-2 Outcomes of HRQL Indices in Participation Models at all Times

Variable	1 month			3 months			6 months			12 months		
	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD
Stroke Impact Scale (SIS)												
Participation	610	52.0	29.3	528	64.9	27.9	477	69.9	27.1	445	75.1	25.9
SIS Recovery	598	60.3	27.3	511	70.0	23.3	470	72.0	22.4	438	76.0	21.9
Euroquol (EQ-5D)												
Usual activities	529	2.0	0.7	460	1.8	0.7	417	1.7	0.7	412	1.6	0.7
EQ-5D VAS	504	66.8	22.1	450	72.1	19.6	400	74.7	18.0	400	76.5	18.4
Short Form 36 (SF-36)												
Role physical	524	23.2	37.1	464	40.5	43.4	411	51.8	43.8	405	62.2	42.6
Role emotional	519	58.0	45.0	462	70.0	40.7	410	80.3	36.0	403	83.0	33.9
Social function	522	48.6	35.1	463	65.8	32.6	411	70.2	31.0	405	75.8	28.5
General health	520	63.7	20.3	458	64.4	19.5	411	66.4	19.9	402	66.4	20.8
Preference Based Stroke Index (PBSI)												
Recreational activities	593	2.0	0.8	528	1.9	0.8	476	1.9	0.8	446	1.8	0.8
Work / activity	592	2.3	0.7	528	2.1	0.7	474	2.1	0.8	447	2.0	0.8
Driving	588	3.0	1.0	529	2.8	1.1	473	2.7	1.2	446	2.6	1.3
SIS, EuroQol VAS, SF-36 higher score = better outcome						EQ5D usual activities and PBSI lower = better						

The measurement model

A model was developed with the four latent variables (social function, role function, restricted roles, and health efficacy) and the measured variables described in Table 8-1. Due to suggestions from the modification indices, an error correlation was added between general health perceptions of the SF-36 and the EuroQol VAS. This is reasonable, as they measure related concepts. The correlation between the social function and role function latent variables was so high at 0.93 that they were likely measuring the same construct. The two latent variables were therefore combined into one latent variable, renamed 'accomplishment'. This led to a model with reasonable fit.

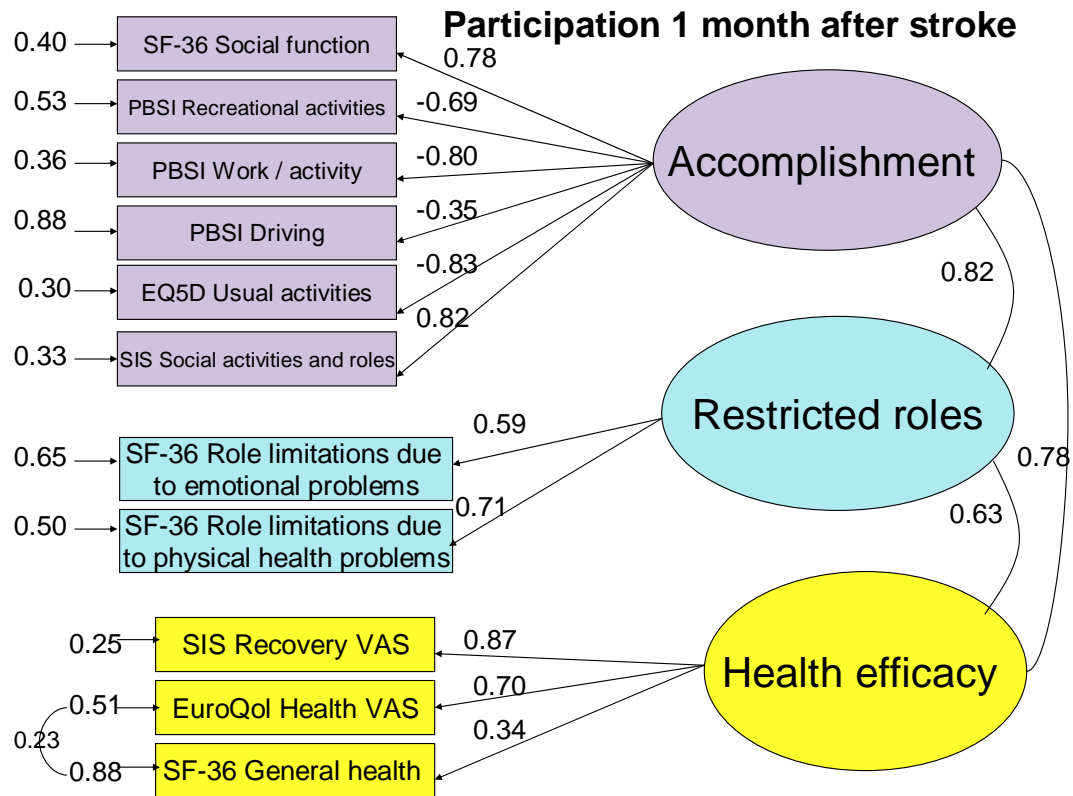
Table 8-3 summarizes the cross sectional participation model at one month, from which all models were built, and 12 months, a time at which participation may be a particular focus for stroke survivors living in the community. Models can not formally be compared cross-sectionally due to the use of full information maximum likelihood (FIML) estimation, as the covariance matrices estimated with FIML are different for each model. Fit statistics are described in detail in Chapter 1. Fit statistics available with FIML include the chi-square (χ^2) and the root mean square error of approximation (RMSEA). The p values for the χ^2 suggest a poor fit, however, having a significant χ^2 is common with large sample sizes¹⁴. It is therefore important to look at other measures of fit as well¹⁴. The normed χ^2 and RMSEA suggest a reasonable fit for both time 1 and time 4. Figure 8-1 displays the participation model at time 1 (1 month). Figure 8-2 displays the participation model at time 4 (12 months).

Table 8-3 Cross-sectional models – Participation

Time	χ^2 (p value)	df	Normed χ^2	RMSEA (90%CI)
1 month	132.8 (0.00)	40	3.3	0.061 (0.050 ; 0.073)
12 months	118.1 (0.00)	40	3.0	0.066 (0.052 ; 0.079)

Normed χ^2 (χ^2 /df) - values up to 5 indicate reasonable fit
 RMSEA \leq 0.05 close fit; 0.05 to 0.08 reasonable fit; \geq 0.10 poor fit¹⁴
 CI = confidence interval, df = degrees of freedom

Figure 8-1 Participation Model – 1 month

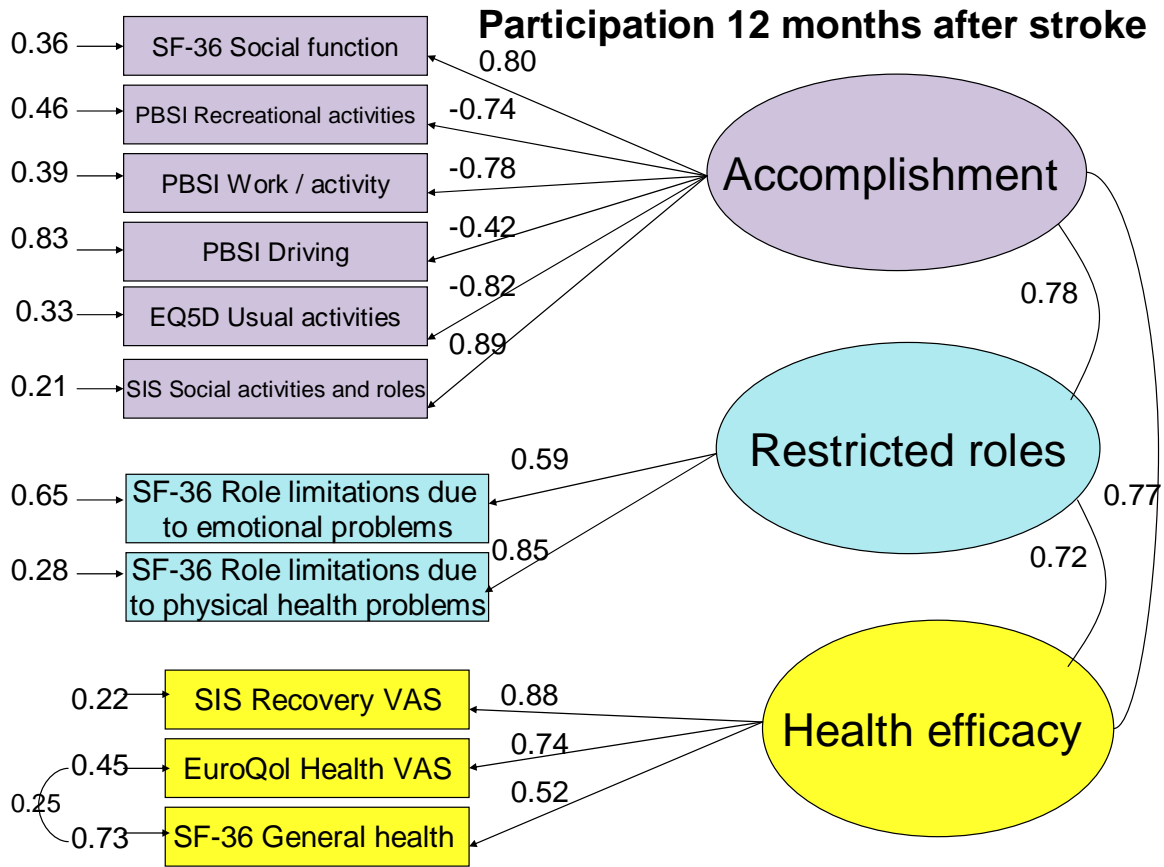


$\chi^2 = 132.83$, $df=40$, $p=0.0000$, $RMSEA = 0.061$ Standardized model shown

Note: Negative correlations reflect the measures PBSI and the EQ5D, where the highest score is the worse health condition.

SF-36 = Short Form 36, SIS = Stroke Impact Scale, PBSI = Preference Based Stroke Index, EQ5D = EuroQol 5 dimensions

Figure 8-2 Participation Model – 12 months



$\chi^2 = 118.13$, $df=40$, $p=0.0000$, $RMSEA = 0.066$ Standardized model shown

Note: Negative correlations reflect the measures PBSI and the EQ5D, where the highest score is the worse health condition.

SF-36 = Short Form 36, SIS = Stroke Impact Scale, PBSI = Preference Based Stroke Index, EQ5D = EuroQol 5 dimensions

Response Shift

After the cross sectional model was developed, all four times were combined into a longitudinal model, including a mean structure. Including the means provides intercepts for the regression equation that represents each path in the model. The longitudinal model had a good fit with RMSEA = 0.034 (90% confidence interval: 0.031 to 0.037), and normed $\chi^2 = 1.8$. The χ^2 fit statistic provided conflicting appraisal of the fit as is common with large samples (χ^2 1359.71 and 766 degrees of freedom (df); $p=0.00$). A 'no response shift' model was developed from the longitudinal model that included constraints of equal intercepts, paths, and error variances across times. If paths, intercepts, and error variances are equal over time, then reconceptualization (path loading pattern), reprioritization (path loading magnitude), and / or recalibration (intercepts or error variances) does not occur.

The χ^2 difference between the longitudinal model and the constrained model was 539.55 with 81 df ($p<0.001$). The constrained model had a significantly worse fit compared to the longitudinal model, suggesting that there is response shift in participation over one year post stroke.

Discussion

It has been proposed that response shift is likely to occur in participation post stroke ⁷. This study supported that assumption, using SEM to identify response shift with a model of participation that included three latent variables: accomplishment, restricted roles, and health efficacy. This is a novel way of modeling response shift in participation over one year post stroke.

Examples of ways in which response shift may occur in participation are as follows: an individual may undergo reprioritization in participation if driving is the most important accomplishment prior to a stroke, but six months after the stroke, he feels that driving is less important and social activities with visiting friends are more important. Recalibration may be demonstrated in an example where at one month post stroke, a woman describes her health as 60/100 on the EuroQol VAS. Twelve months later, she has returned home and has been able to resume her volunteer work, realizing that her health is now 90/100, but was likely only 50/100 at one month, rather than 60.

Physical therapists and other rehabilitation professionals have training to be able to assist clients post stroke in increasing their abilities and decreasing barriers to participation.

They are also in a position to assist the client in redefining expectations of what optimal participation entails⁷. This is one way in which response shift may be the goal of a rehabilitation intervention. Other examples of response shift as a focus of treatment have been described¹⁵⁻¹⁷.

In a previous study, a variety of factors including upper and lower extremity motor coordination and upper extremity abilities predicted long term participation at 2-4 years post stroke³. The authors of that study pointed out that except for age, all of the predictors could be addressed with treatment. This supports the concept of the inclusion of recovery from the client's perspective (which may be influenced by treatment) as an aspect of participation.

As in the mental health and physical function models, the sample size of 677 may be considered too small. The number of cases to free parameters estimated should be at least

10:1¹⁴. With the large longitudinal model used in this study, there are over 100 parameters estimated; the sample size may therefore be small for the longitudinal model.

This study will be continued to identify the localization of response shift. The timing and type of response shift present will be determined, as well as which measured variables are affected. It would also be beneficial to test the model of participation with other groups of clients post stroke and clients with other chronic health conditions.

Conclusion

Response shift was identified in the participation model over a period of one year post stroke. The findings in this study can serve to inform future research of response shift and models of participation post stroke. This is a novel description of post stroke response shift in a longitudinal model of participation. Consideration of the assessment and training of response shift in participation post stroke needs to be considered.

Reference List

- (1) Sprangers MA. Quality-of-life assessment in oncology. Achievements and challenges. *Acta oncologica Stockholm, Sweden* 2002;41(3):229-37.
- (2) Rochette A, Desrosiers J, Noreau L. Association Between Personal and Environmental Factors and the Occurrence of Handicap Situations Following a Stroke. *Disability and Rehabilitation* 2001;23(13):559-69.
- (3) Desrosiers J, Noreau L, Rochette A, Bourbonnais D, Bravo G, Bourget A. Predictors of Long-term Participation After Stroke. *Disability and Rehabilitation* 2006;28(4):221-30.
- (4) Sveen U, Thommessen B, Bautz-Holter E, Wyller T, Laake K. Well-being and Instrumental Activities of Daily Living After Stroke. *Clinical Rehabilitation* 2004;18:267-74.
- (5) Mayo N, Wood-Dauphinee S, Cote R, Durcan L, Carlton J. Activity, Participation, and Quality of Life 6 Months Poststroke. *Archives of Physical Medicine and Rehabilitation* 2002;83:1035-42.
- (6) World Health Organization. Towards a Common Language for Functioning, Disability and Health - ICF. *WHO website* 2002; Available at: URL: www3.who.int/icf/beginners/bg.pdf. Accessed April 12, 2006.
- (7) Rochette A, Korner-Bitensky N, Levasseur M. 'Optimal' Participation: A Reflective Look. *Disability and Rehabilitation* 2006;28(19):1231-5.

- (8) Han B, Small B, Haley W. The Structure of Self-Rated Health Among Community-Dwelling Older Adults with Stroke. *Home Health Care Services Quarterly* 2001;20(4):1-15.
- (9) Wood-Dauphinee S. Quality of Life Post Stroke. *Brainwaves - newsletter for members of the Canadian Stroke Network* 2003;Issue 19(May):2-4.
- (10) SAS Institute. *SAS/STAT User's Guide, Version 9*. Cary,NC: SAS Institute, Inc.; 2004.
- (11) Jöreskog K, Sörbom D. *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Lincolnwood, IL: Scientific Software International Inc.; 1993.
- (12) Oort FJ. Using Structural Equation Modeling to Detect Response Shifts and True Change. *Quality of Life Research* 2005;14:587-98.
- (13) Stewart A. The Medical Outcomes Study Framework of Health Indicators. In: Stewart A, Ware J, editors. *Measuring Functioning and Well-Being*. Durham: Duke University Press; 1992. p. 12-24.
- (14) Kline R. *Principles and Practice of Structural Equation Modeling*. 2 ed. New York: The Guilford Press; 2005.
- (15) Wilson IB. Clinical Understanding and Clinical Implications of Response Shift. *Social Science and Medicine* 1999;48:1577-88.

- (16) Schwartz C, Andresen EM, Nosek M, Krahn G, RRTC Expert Panel on Health Status Measurement. Response Shift Theory: Important Implications for Measuring Quality of Life in People With Disability. *Archives of Physical Medicine and Rehabilitation* 2007;88:529-36.
- (17) Osborne R, Hawkins M, Sprangers M. Change of Perspective: A Measureable and Desired Outcome of Chronic Disease Self-Management Intervention Programs That Violates the Premise of Preintervention / Postintervention Assessment. *Arthritis and Rheumatism* 2006;55(3):458-65.

Chapter 9: Conclusion

“Although the presence of response shift is intuitive, it is a very abstract concept that is difficult to capture”¹.

The overall objective of this thesis was to assess response shift in construct specific health related quality of life (HRQL) models. Structural equation modeling (SEM) was used to identify response shift post stroke over four time points in mental health, physical function, and participation. The evaluation of each HRQL construct used different SEM methodology, different measurement models, and expanded time-frames compared to a previous study which did not identify response shift post stroke with SEM². The previous study used the SF-36 as a generic index of HRQL; this study evaluated HRQL constructs separately, providing construct-specific information. The available timeframe of 1, 3, 6, and 12 months post stroke allowed for a broader evaluation of the potential timing of response shift. The methodology used allowed for the evaluation of all types of response shift, which had not been previously possible.

Summary of Hypotheses

Hypothesis 1:

a. ‘The factor structure of the latent construct mental health is not invariant over time’. The factor loading patterns and factor loading values were invariant over time (did not change) in the mental health model – reconceptualization and reprioritization were not identified. Non-uniform recalibration was identified, however; the error variances of 4 measured variables were not invariant over time (did change).

b. ‘The mean pattern of the latent domain mental health is not invariant over time’. The values of the intercepts of 2 measured variables were not invariant over time (did change) in the mental health model – uniform recalibration was therefore identified.

c. ‘Response shift will be detected’. Response shift in the form of uniform recalibration was noted in the SF-36 role emotional and PBSI self esteem subscales around 12 months. Non-uniform recalibration was identified in the EQ5D anxiety/depression, SIS emotional, and SF-36 mental health subscales around 12 months, and around the 6 month time in SF-36 mental health.

Hypothesis 2:

a. ‘The factor structure of the latent construct physical function is invariant over time’.

The factor loading patterns were invariant (did not change), but the factor loading values were not invariant over time (did change) in the physical function model –reprioritization was identified in 1 measured variable. Non-uniform recalibration was also identified in 1 measured variable.

b. ‘The mean pattern of the latent domain physical function is invariant over time’.

The values of the intercepts of 3 measured variables were not invariant over time (did change) in the physical function model – this reflects uniform recalibration.

c. ‘Response shift will not be detected’.

The detection of response shift was unexpected in the physical function construct. We assumed that the reason may be in part due to the wording of the questions, such as the

use of the word ‘difficulty’. To summarize, both reprioritization and uniform recalibration were identified in PBSI physical activities at time 4. Uniform recalibration was also identified in PBSI stairs at time 4 and SIS hand function at time 2. Non-uniform recalibration was identified a time 4 in the PBSI walking item. The SIS hand function subscale and the PBSI walking item both utilized the term ‘difficulty’.

Hypothesis 3:

- a. ‘The factor structure of the latent construct participation is not invariant over time’.
- b. ‘The mean pattern of the latent domain participation is not invariant over time’.
- c. ‘Response shift will be detected’.

Response shift was detected in the overall participation model. Identification of the localization of the type and timing of response shift will continue.

Hypothesis 4:

‘Response shift will be seen at 12 months post stroke’.

With the identification of response shift in physical function, it was clear that response shift was identified in 4 measured variables at time 4, and once at time 2. Similarly with the mental health model, response shift was identified in 5 measured variables at time 4 and once at time 3. Therefore, response shift appears to occur primarily at a later time (time 4 / 12 months) rather than earlier after a stroke.

Overall Summary

This study developed and applied a framework to guide researchers in using structural equation modeling (SEM) to identify response shift when there are greater than two time points (Chapter 5). There are many possible localized sources of response shift when evaluating multiple time points, such as timing of response shift, type of response shift, or response shift in particular measured variables. The framework also addresses other modeling issues such as model validation, familywise Type I error and considering the purpose of the research. The purpose of the research may be exploratory to identify localization of response shift to inform and guide future studies or theory-driven to confirm previous study results.

This was the first time that models of individual constructs of HRQL (physical function, mental health, and participation) were utilized to identify response shift post stroke. This study provides information about the presence of response shift in HRQL up to 12 months post stroke. The demonstration of models of physical function, mental health and participation post stroke provide insight into the constructs and will be beneficial in determining future models in other samples of individuals post stroke. There were some similarities between the latent variables in the physical function model and previous models of physical function in individuals with other chronic health conditions^{3, 4}.

To our knowledge, this is the first account of a longitudinal model of physical function post stroke as well as the first description of response shift in physical function and participation. With response shift present in physical function, the ways which we evaluate physical function in the clinic and in research studies, the type of measures that we use, and the way in which change in physical function is analyzed will be affected.

In the physical function model, both reprioritization and uniform recalibration were identified in the PBSI physical activities item at time 4 (12 months). This is not surprising, as the various types of response shift are likely interdependent; they may occur together, in parallel, or at the same rate⁵. Recalibration may therefore occur along with reconceptualization or reprioritization because the meanings of the anchors of a measurement scale may change if the response scale increases or decreases⁵. It has not, however, yet been established if reconceptualization, reprioritization, and recalibration occur independently⁶.

It has been suggested in the past that due to the necessity for more research in response shift, “it may be reasonable in some instances to limit one’s investigation to simply detecting response shift”⁷. Using the framework that we developed to identify response shift with greater than 2 time points, however, this would mean stopping at step 1 – determining that there is response shift somewhere in the model. What, therefore, is the benefit of identifying the localization (timing, type, measure) of response shift?

The benefits of identifying the timing, type, and measured variables associated with response shift

Timing

For interventions aimed at inducing ‘positive’ response shift, it would be beneficial to be aware of the timeframe at which response shift is most likely to occur. An example of an intervention and outcome measure aimed at inducing and identifying response shift is described by Osborne et al.⁸. Of the 11 locations where response shift was identified across the physical function and mental health models using the exploratory approach described in our framework, response shift was identified in 9 locations at time 4, once at

time 3, and once at time 2. Response shift seems to occur primarily at a later time post stroke – identified at the 12 month period. Perhaps the catalyst for response shift is returning home to the community- this needs further exploration. Response shift may even occur later than 12 months post stroke; this also needs further evaluation. See Table 9-1 for a summary of response shift localization in the mental health and physical function models.

Table 9-1 Response Shift Localization

Mental health			Physical function			
months	recalibration		months	reprioritization	recalibration	
	uniform	non-uniform			uniform	non-uniform
12	SF-36 role emotional		12	PBSI physical activities		
	PBSI self esteem				PBSI activities	
		SIS emotional			PBSI stairs	
		EQ5D anxiety/depression				PBSI walking
		SF-36 mental health	3		SIS hand function	
6		SF-36 mental health				

In previous work on response shift post stroke, recalibration was identified between 6 weeks and 6 months in the EuroQol VAS with the then test ⁹. Reconceptualization and reprioritization were also identified with the PGI between 6 weeks and 6 months ¹⁰. A previous study with SEM, however, did not identify response shift at the same time-frame ². The studies did not have data available at 12 months, and the SEM study used a

different SEM methodology, therefore, it is difficult to compare studies. It does appear, however, in the current study that response shift does occur around the one year timeframe post stroke. This has possible future implications for treatment interventions in the community which are based on improving activity limitations and participation in the ‘chronic’ stroke survivor population. True change as well as response shift should be evaluated in both clinical and research projects with this population.

It has been suggested that recalibration is more likely in the first few months after a threatening event and that clients with more severe symptoms engage in recalibration for longer time intervals than those with milder symptoms¹¹. It is unknown if the individuals in the current study who underwent response shift around the one year time had more severe strokes, due to the group level analysis. This would be interesting to evaluate in future research studies. A multi-group analysis in SEM comparing those with more and less severe activity limitations is a future possibility.

When using the framework, the choice between the exploratory and theory-driven approaches should be made early on. It was demonstrated that response shift was identified at time 2 in the mental health model with the theory-driven approach, but not in the exploratory method due to correcting for familywise Type I error due to multiple testing. Results between the two methods can vary, due to the number of models tested. The choice of methods used must be based on evidence available in the area of study.

Type of response shift

In mental health, recalibration (uniform and non-uniform) was identified six times. In physical function, reprioritization was identified in one measured variable, and

recalibration (non-uniform and uniform) was identified in 4 measured variables. Therefore, the most predominant type of response shift identified across the two constructs of physical and mental health was recalibration. In a clinical setting, it may be appropriate to measure HRQL in individuals with a method that would identify recalibration specifically, such as the then test.

The framework that was developed allows for identification of specific types of response shift based on theory and past research (theory-driven approach). It is suggested, however, that until more is known about response shift, that research with new models focus primarily on the exploratory approach, to generate information for future studies. Future research comparing the two SEM methods of identifying response shift localization may be beneficial, to determine if similar types of response shift are identified^{15, 16}. In the method described by Schmitt, however, only reconceptualization and recalibration are evaluated.

Measured Variable

Knowing which measures and items demonstrate response shift may give us some information about items in a measure being likely to identify response shift. In the mental health model, there were 6 instances of response shift identified: three in the SF-36, one in the PBSI, one in the EQ5D, and one in the SIS. In the physical function model, five instances were identified, four in the PBSI and one in the SIS. It appears that the PBSI may be particularly susceptible to response shift – this deserves further research.

Response shift may influence the measurement characteristics of outcome measurement

tools, so it is important to measure response shift to assist in understanding assessment of HRQL¹².

Response shift was identified in the physical function model, where it had not been hypothesized to occur. When it was apparent from Step 1 of the framework that response shift was present somewhere in physical function, we made a secondary hypothesis prior to identifying the localization of response shift that perhaps the wording of some of the measured variables might be more susceptible to response shift. The term ‘difficulty’ is a response option that provides an evaluation-based response by the individual. It has been suggested that changes in scores of measurement scales over time that are not considered to be true change may be interpreted as response shift in evaluation-based (‘how difficult is it to...’) methods, while it may be measurement error in performance based (such as a timed walking test) measures¹³. The evaluation of ‘difficulty’ may partly explain why response shift occurred in the model of physical function. It can not, however, be the only explanation; two of the four measured variables in which response shift was suggested contained the evaluation of ‘difficulty’. Previous research with older women has found that some individuals continue to be physically active despite described difficulties while others are inactive despite few difficulties¹⁴. This concept needs further evaluation.

In an evaluation of the effect size of measured variables that demonstrated response shift, the PBSI physical function variable had a small true effect size of -0.35, but when the response shift effect size was included, the overall observed effect size was moderate at -0.50. Including response shift appears to change the interpretation of the amount of change that occurred, making it an important phenomenon to measure. See Table 9-2 for a summary of future research based on this thesis work.

Table 9-2 Future research based on thesis work

General	Framework	Physical Function Model	Participation Model
Compare the methods described by Oort and Schmitt with the same sample ^{15, 16}	Utilize other validation methods, such as a comparison of the model in question to other populations	Utilize both performance based measures and self-report measures	Determine localization of response shift in participation
Study the possibility that the catalyst for response shift may be return to the community	Validation of the identification of the localization of response shift steps as well as validating the full model	Comparison of models of physical function with only performance based measures and models with only evaluation-based “difficulty” measures would be of interest	Evaluate the participation model in another sample of community dwelling individuals post stroke
Evaluation of response shift at longer time (≥ 18 months) post stroke	Adapt to other statistical methods i.e. SEM method described by Schmitt.	Identify response shift in physical function in other samples of individuals post stroke and in other chronic condition populations	Evaluate the participation model in individuals with other chronic health conditions
Evaluate if severity of activity limitations has affect on type and timing of response shift with multi-group analysis	Investigate other types of correction for familywise Type I error	Explore response shift in pain post stroke, given estimates of the frequency of shoulder pain and pain in other areas of the body	
Investigate the susceptibility of the PBSI to response shift			
Further evaluation of the clinical importance of response shift			
Compare methods with convergent validity studies			

Overall limitations

There are some limitations to the framework. The framework does not identify an order of testing of response shift types. We did however, test in the following order: reconceptualization, reprioritization, recalibration (uniform, non-uniform). This order reflects the order suggested by Donaldson, as described in Chapter 1 (assessment of measurement invariance). Factor loading invariance is considered first in classical factor analysis; if the unstandardized regression coefficients are the same across populations or time, it is assumed that the models are the same¹⁷, or in the case of response shift, no reconceptualization (change in factor loading pattern) or reprioritization (change in factor loading values) has occurred. However, it has been questioned that if the sequence of testing does not alter the end result, order of testing may not matter⁶.

In the framework, validation of the full model only is suggested. There is opportunity to assess validation through the process of the identification of location of response shift. We did also take an initial look at comparing FIML and robust statistics with multiple imputation and found similar, but not identical identification of the location of response shift in the mental health model. This deserves future study.

Power and sample size in large longitudinal models needs to be considered. In the mental health, physical function, and participation models, the sample size was 677. There were more than 100 parameters estimated in each of the four time unconstrained models. It has been suggested that the number of cases to free parameters estimated be at least ten to one¹⁸. For the unconstrained longitudinal model, a sample size of 677 might be considered too small. Clearly, sample size requirements increase when the model is

complex due to the inclusion of multiple time points, latent variables or measured variables.

Structural equation modeling involves group level analysis, so that results are averaged across all individuals. Not all individuals would be expected to undergo response shift and those who do experience response shift may not all experience it at the same time¹⁹. This was demonstrated by Mayo et al who used this same data set and a method which used the residuals generated from a growth model where measured health behaviours or functions were used to predict perceived health¹⁹. They found that there were subgroups of the population who did show a response shift (positive and negative) but the majority of subjects (67%) showed no response shift¹⁹.

It has been stated that “response shift detection with the SEM approach relies on the choice of the measurement model; the findings have no implications beyond the model”²⁰. Therefore, all of the models should be evaluated on different samples of individuals post stroke and also in other chronic health conditions.

Implications

Measurement of change in HRQL will be inaccurate if response shift is present, but not accounted for. Response shift must therefore be taken into account. Treatment effects can be underestimated if response shift is not considered when evaluating response shift susceptible constructs of HRQL. It has been demonstrated that change in HRQL can be underestimated when recalibration occurs, as measured by the then test in clients post stroke⁹. The effect size of HRQL change can be estimated, taking response shift into account, after evaluation with SEM: HRQL change is often larger when response shift is

accounted for ²¹. This was demonstrated in the physical function model, where observed change (true change plus response shift) in physical activity had a moderate effect size which became small when response shift was removed. Future response shift studies of the clinical importance of response shift are required.

Increased knowledge of response shift will therefore affect the way in which HRQL measures are used in clinical, research, and policy decisions. Clinically, if recalibration response shift is suspected in a particular construct of HRQL (as demonstrated in mental health and physical function), a design method such as a then test could be used to capture response shift over time during treatment. It is important to measure response shift in clinical trials research, as the estimates of the treatment effects may be underestimated and therefore inaccurate, possibly leading to a conclusion that is false negative ¹². Estimation of HRQL in studies of health services delivery should take response shift into account to be able to correctly interpret the results, leading to changes in health policy ²². There is no gold standard for the measurement of response shift. Researchers should aim to confirm results by comparing methods (i.e., statistical and design methods) in convergent validity studies, as suggested in the first manuscript.

Response shift is often the focus of medical treatments, including rehabilitation ^{8, 23, 24}. For example, a health professional may want an individual to have a ‘good outlook’ (good HRQL) despite limitations in physical function, with a focus on the positive. In a self-management program, the goal is often to induce response shift ⁸. Rehabilitation interventions with individuals may teach response shift (such as in participation post stroke) ²⁵, therefore response shift assessment should be part of the self report measures

used to measure treatment effectiveness²⁴. When the goal of an intervention is response shift, it needs to be specifically measured.

If response shift is thought to be present, as in this demonstration, the calculation of changes over time will be inaccurate. In research studies where response shift is suspected and design methods are not possible, then statistical methods should be utilized. The framework presented in Chapter 5 may be helpful in further structuring the use of SEM to identify the location of response shift.

Reference List

- (1) Ahmed S, Mayo N. Author's Response: Using Structural Equation Modeling to Evaluate Response Shift. *Journal of clinical epidemiology* 2007;60:427-8.
- (2) Ahmed S, Mayo N, Corbiere M, Wood-Dauphinee S, Hanley J, Cohen R. Change in Quality of Life of People with Stroke Over Time: True Change or Response Shift? *Quality of Life Research* 2005;14:611-27.
- (3) Sousa K, Kwok O. Putting Wilson and Cleary to the Test: Analysis of a HRQOL Conceptual Model Using Structural Equation Modeling. *Quality of Life Research* 2006;15:725-37.
- (4) Wang P, Badley E, Gignac M. Activity Limitation, Coping Efficacy and Self-Perceived Physical Independence in People with Disability. *Disability and Rehabilitation* 2004;26(13):785-93.
- (5) Schwartz C, Sprangers M. Discussion: Theoretical Reflections. In: Schwartz C, Sprangers M, editors. *Adaptation to Changing Health - Response Shift in Quality-of-Life Research*. Washington: American Psychological Association; 2000. p. 73-7.
- (6) Ahmed S, Mayo N. Response to Donaldson's Commentary. *Quality of Life Research* 2005;14:2357-8.
- (7) Brossart D, Clay D, Willson V. Methodological and Statistical Considerations for Threats to Internal Validity in Pediatric Outcome Data: Response Shift in Self-Report Outcomes. *Journal of pediatric psychology* 2002;27(1):97-107.

- (8) Osborne R, Hawkins M, Sprangers M. Change of Perspective: A Measureable and Desired Outcome of Chronic Disease Self-Management Intervention Programs That Violates the Premise of Preintervention / Postintervention Assessment. *Arthritis and Rheumatism* 2006;55(3):458-65.
- (9) Ahmed S, Mayo N, Wood-Dauphinee S, Hanley J, Cohen S. Response Shift Influenced Estimates of Change in Health-Related Quality of Life Poststroke. *Journal of clinical epidemiology* 2004;57:561-70.
- (10) Ahmed S, Mayo N, Corbiere M, Wood-Dauphinee S, Hanley J, Cohen R. Using the Patient Generated Index to Evaluate Response Shift Post Stroke. *Quality of Life Research* 2005;14:2247-57.
- (11) Schwartz CE, Sprangers MAG, Carey A, Reed G. Exploring response shift in longitudinal data. *Psychology and Health* 2004 February;Vol 19(1):51-69.
- (12) Schwartz C, Sprangers M, Fayers P. Response Shift: You Know it's There But How do You Capture it? Challenges for the Next Phase of Research. In: Fayers P, Hay R, editors. *Assessing Quality of Life in Clinical Trials - Methods and Practice*. 2 ed. Oxford University Press; 2005. p. 275-90.
- (13) Schwartz CE, Rapkin BD. Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. *Health and quality of life outcomes [electronic resource]* 2004 March 23;2(1):16.
- (14) Rantanen T, Guralnik J, Sakari-Rantala R, Leveille S, Simonsick E, Ling S, Fried L. Disability, Physical Activity, and Muscle Strength in Older Women: The

Women's Health and Aging Study. *Archives of Physical Medicine and Rehabilitation* 1999;80:130-5.

- (15) Oort FJ. Using Structural Equation Modeling to Detect Response Shifts and True Change. *Quality of Life Research* 2005;14:587-98.
- (16) Schmitt N. The Use of Analysis of Covariance Structures to Assess Beta and Gamma Change. *Multivariate Behavioural Research* 1982;17:343-58.
- (17) Donaldson G. Structural Equation Models for Quality of Life Response Shifts: Promises and Pitfalls. *Quality of Life Research* 2005;14:2345-51.
- (18) Hays R, Revicki D, Coyne K. Application of Structural Equation Modeling to Health Outcomes Research. *Evaluation and the Health Professions* 2005;28(3):295-309.
- (19) Mayo N, Scott S, Dendukuri N, Ahmed S, Wood-Dauphinee S. Identifying Response Shift Statistically at the Individual Level. *Quality of Life Research - accepted* 2008.
- (20) Oort F. Towards a Formal Definition of Response Shift (In Reply to G.W. Donaldson). *Quality of Life Research* 2005;14:2353-5.
- (21) Oort FJ, Visser M, Sprangers MA. An Application of Structural Equation Modeling to Detect Response Shifts and True Change in Quality of Life Data from Cancer Patients Undergoing Invasive Surgery. *Quality of Life Research* 2005;14:599-609.

- (22) Ahmed S, Mayo N, Wood-Dauphinee S, Hanley J, Cohen S. The Structural Equation Modeling Technique Did Not Show a Response Shift, Contrary to the Results of the Then Test and the Individualized Approaches. *Journal of clinical epidemiology* 2005;58:1125-33.
- (23) Wilson IB. Clinical Understanding and Clinical Implications of Response Shift. *Social Science and Medicine* 1999;48:1577-88.
- (24) Schwartz C, Andresen EM, Nosek M, Krahn G, RRTC Expert Panel on Health Status Measurement. Response Shift Theory: Important Implications for Measuring Quality of Life in People With Disability. *Archives of Physical Medicine and Rehabilitation* 2007;88:529-36.
- (25) Rochette A, Korner-Bitensky N, Levasseur M. 'Optimal' Participation: A Reflective Look. *Disability and Rehabilitation* 2006;28(19):1231-5.