

Measurement of Health-Related Quality of Life in Canadians with
Neurological Conditions:
A Comparison of the SF6D and HUI3

by

Hannah Abel

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
March 2014

© Copyright by Hannah Abel, 2014

DEDICATION

To my daughters, Cleo and Carmen, who are a source of joy and pride and a reminder of the important things in life...

TABLE OF CONTENTS

LIST OF TABLES.....	v
LIST OF FIGURES.....	vi
ABSTRACT.....	vii
LIST OF ABBREVIATIONS AND SYMBOLS USED.....	viii
ACKNOWLEDGEMENTS.....	ix
CHAPTER 1. INTRODUCTION.....	1
CHAPTER 2. BACKGROUND.....	4
2.1 PREFERENCE-BASED HRQoL INSTRUMENT USE.....	4
2.2 DIFFERENTIAL MEASUREMENT OF HRQoL.....	5
2.3 DESCRIPTIVE HEALTH CLASSIFICATION SYSTEMS OF THE SF6D AND HUI3.....	6
2.3.1 Operationalization of HRQoL.....	6
2.3.2 Instrument content and structure.....	6
2.3.3 Contextual framing differences.....	7
2.3.4 Comprehensiveness of the SF6D and HUI3 descriptive systems.....	8
2.4 SCORING SYSTEMS OF THE SF6D AND HUI3.....	8
2.5 INSTRUMENT DIFFERENCES AND IMPLICATIONS FOR HRQoL MEASUREMENT.....	9
2.6 AGREEMENT BETWEEN THE SF6D AND HUI3.....	10
2.7 FACTORS AFFECTING UTILITY DIFFERENCES BETWEEN THE SF6D AND HUI3.....	11
2.8 IMPLICATIONS OF MEASUREMENT DISCORDANCE.....	12
2.9 LIMITATIONS IN THE LITERATURE.....	13
2.10 SUMMARY.....	14
2.11 PURPOSE OF THE STUDY.....	14
2.12 OBJECTIVES.....	15
CHAPTER 3. SENSITIVITY IN MEASURING HEALTH-RELATED QUALITY OF LIFE IN CANADIANS WITH NEUROLOGICAL CONDITIONS: A COMPARISON OF THE SF6D AND HUI3.....	17
ABSTRACT.....	18
3.1 INTRODUCTION.....	19
3.2 METHODS.....	21
3.2.1 Sample and data collection.....	21
3.2.2 Measures of HRQoL.....	22

3.2.3 <i>Statistical Analysis</i>	25
3.3 RESULTS	26
3.4 DISCUSSION	28
3.5 CONCLUSIONS	32
CHAPTER 4. DISCORDANCE IN UTILITY MEASUREMENT IN CANADIANS WITH NEUROLOGICAL CONDITIONS: A COMPARISON OF THE SF6D AND HUI3	38
ABSTRACT	39
4.1 INTRODUCTION	40
4.2 METHODS	42
4.2.1 <i>Sample and data collection</i>	42
4.2.2 <i>Measures of HRQoL</i>	43
4.2.3 <i>Statistical Analysis</i>	46
4.2.4 <i>Analysis Steps</i>	48
4.3 RESULTS	52
4.4 DISCUSSION	55
4.5 CONCLUSIONS	61
CHAPTER 5. CONCLUSION.....	73
5.1 LIMITATIONS.....	74
5.2 IMPLICATIONS FOR FUTURE RESEARCH.....	75
REFERENCES	77
APPENDIX A: HEALTH UTILITY INDEX - MARK 3 (HUI3).....	86
APPENDIX B: SF-6D UTILITY INDEX (QUALITY METRIC).....	90
APPENDIX C: QUALITY OF LIFE IN NEUROLOGICAL DISORDERS	93
APPENDIX D: MEASURING AGREEMENT BETWEEN INSTRUMENTS.....	95
APPENDIX E: FITTED VALUES OF DUTY BY DIFFERENT REGRESSION TECHNIQUES.....	97
APPENDIX F: GRAPHICAL REPRESENTATION OF REGRESSION MODEL USING 3 LINEAR SPLINES	98
APPENDIX G: DISTRIBUTION OF MEAN UTILITY	99

LIST OF TABLES

Table 2.1	Comparison of the HUI3 & SF6D by descriptive health classification system and scoring system characteristics..	16
Table 3.1	Sociodemographic and clinical characteristics of the study population.....	33
Table 3.2	Quality of Life of the study population	34
Table 3.3	HUI3 attribute-to-attribute correlations with Neuro-QoL Scale dimensions (Spearman’s Rho).....	35
Table 3.4	SF6D attribute-to-attribute correlations with Neuro-QoL Scale dimensions (Spearman’s Rho).....	36
Table 3.5	Mean utility scores by normal, mild and moderate to severe impairment of Neuro-QoL domains	37
Table 4.1	Characteristics of the study population.....	65
Table 4.2	Frequency and utility of neurological conditions included in the sample.....	66
Table 4.3	Spearman’s Rho Correlation Matrix for SF6D and HUI3 HRQoL domains.....	67
Table 4.4	Distribution of participant responses (%) across the SF6D and HUI3 health domain levels.....	68
Table 4.5	Multiple regression modeling associations between participant characteristic factors and utility difference.....	69
Table 4.6	Multiple regression modeling associations between neurological condition type and utility difference.....	70
Table 4.7	Multiple regression modeling associations between neurological condition symptoms and impairments (Neruo-QoL scale variables) and utility difference.....	71
Table 4.8	Distribution of participant responses (%) in SF6D and HUI3 health domain levels in individuals with an HUI3 score of 0.90 or greater.....	72

LIST OF FIGURES

Figure 4.1	Distributions of the SF-6D and HUI3	62
Figure 4.2	Scatter plot of the SF6D and HUI3 with regression line.	63
Figure 4.3	Bland-Altman plot of the SF-6D and HUI3.....	64

ABSTRACT

The objective of this study was to contribute evidence regarding the use of the SF6D and HUI3 in persons with neurological conditions. The data of 776 individuals from the LINC Study was analyzed. The mean utility score of the HUI3 was 0.47 (95% CI 0.45, 0.49) and SF6D was 0.62 (95% CI 0.62, 0.63). Even though the SF6D and HUI3 were sensitive to a variety of HRQoL domains relevant to persons with neurological conditions, they showed only marginal agreement (ICC of 0.41) with a mean utility difference of 0.15 (95% CI 0.13, 0.17). Discordance varied systematically with HRQoL status and was consistent regardless of the participant or impairment characteristics present. Despite sharing a common purpose, the substantial and clinically important differences found between the SF6D and HUI3 cast doubt on whether the utility estimates produced by these instruments are directly comparable or universally valid.

LIST OF ABBREVIATIONS AND SYMBOLS USED

ANOVA	Analysis of Variance
CADTH	Canadian Agency for Drug and Technologies in Health
CI	Confidence Interval
dUTY	Difference in Utility
EQ-5D	Euro-QoL 5D Instrument
HRQoL	Health Related Quality of Life
HUI2	Health Utilities Index – Mark 2 Instrument
HUI3	Health Utilities Index – Mark 3 Instrument
ICC	Intra-class Correlation
LINC	Living with the Impact of a Neurological Condition (Study)
MID	Minimally Important Difference
MS	Multiple Sclerosis
Neuro-QoL	Quality of Life in Neurological Disorders Measures (Short Form Instrument)
NICE	National Institute for Health and Clinical Excellence
NHCC	Neurological Health Charities Canada
OLS	Ordinary Least Squares Regression
QALY	Quality Adjusted Life Year
QWB	Quality of Life and Well-being Instrument
SD	Standard Deviation
SF6D	Short-Form 6D Instrument
SF-36v2	Short-Form 36 (Medical Outcomes Trust and Quality Metric Instrument)
SG	Standard Gamble
SLNCC	Survey of Living with a Neurological Condition in Canada (Statistics Canada)
TTO	Time Trade-Off
VAS	Visual Analogue Scale

ACKNOWLEDGEMENTS

I would like to thank the many individuals who provided encouragement, time, knowledge and support during my Master's program.

To my Thesis committee, Dr. George Kephart, Dr. Tanya Packer, and Dr. Grace Warner, my very great appreciation for their guidance, enthusiastic encouragement, and advice. It was a pleasure to work with such sage and patient mentors. I would like to add a special thanks to my primary supervisor, Dr. George Kephart, for his ongoing confidence in me and in ensuring that I did not get lost in the "trees".

I would also like to extend a heartfelt thanks to my parents, John and Malena, from whom I've learned that hard work and perseverance do payoff and for catching all my misplaced commas and semi-colons over the years.

Last, but definitely not least, thank you to my beloved husband Ross and daughters Carmen and Cleo for their understanding and allowing me time to research and write.

CHAPTER 1. INTRODUCTION

Neurological diseases, disorders and injuries are a significant cause of disability and disease burden in Canada. (1) Collectively, neurological conditions are characterized by diverse symptoms and functional limitations that pose daily challenges to individuals and their families. (1,2) Often the impact of a condition extends beyond physical symptoms and impairments and impacts social relationships and roles. (3-5) Very few neurological conditions are curable. As a result, the emphasis is often on managing the impact of symptoms and the maintenance of quality of life and life satisfaction. (6) Among the options to quantify the “impact” of a neurological condition on everyday life is the measurement of health-related quality of life (HRQoL). (1,7-9)

Health-related quality of life is a multidimensional construct that describes improvements or detriments in quality of life that are central to health such as the physical, mental, and social-emotional aspects of health. Health related quality of life is an important outcome in neurological conditions as it contributes a more comprehensive understanding of disease effects than measures of clinical change alone. Successful evaluation of neurological conditions requires instruments that are sensitive to the functional impairments of the condition as well as capturing aspects of health that are more subjective and personal in nature; to describe and quantify the impact and consequences of a disease on daily life. (6,10,11)

There are a variety of condition-specific and generic HRQoL instruments available to evaluate HRQoL; each has benefits and limitations. Condition-specific instruments can provide an accurate assessment of HRQoL domains particularly relevant to a specific diagnosis. In comparison, generic instruments are purposefully broad in their evaluation of HRQoL to enable comparisons across a range of populations, conditions, and interventions.

Preference-based health related quality of life instruments (also commonly known as multi-attribute / utility /index - instruments) are a unique class of generic HRQoL instruments that evaluate HRQoL in terms of utility. (12-15) “Utility” is an economic concept that represents a preference for the item considered; in the context of HRQoL,

utility represents a preference for health states as “better” or “worse”. Utility instruments use a scoring formula (algorithm) to derive a utility score based on an individual’s self-reported responses within the instruments questionnaire. The scoring algorithm applies different weights to health states based on pre-determined community-derived preferences for the health states. Although individual’s completing these questionnaires provide self-reports of HRQoL – the resulting preferences attached to the utility scores generated are not that of the respondent, but rather that of an average member of the community.

Utility instruments have gained widespread use in the past two decades to evaluate and monitor population health; inform clinical decision-making; evaluate programs and interventions; and are recommended for use in economic evaluations (e.g. by the Canadian Agency for Drugs and Technologies in Health (CADTH) and the National Institute for Health and Clinical Excellence (NICE)). (8,12,14,16-20) A limited number of utility instruments currently dominate the literature and include: the Health Utilities Index – Mark 2 (HUI2) (21) and Mark 3 (HUI3), (22,23) the EuroQol (EQ-5D), (24,25) Quality of Life and Well-being (QWB) (26) and Short-Form 6D (SF6D). (26)

The appeal of utility instruments, over other measures of HRQoL, is their ability to provide an overall assessment of health status and quality of life in a single summary score that simultaneously weighs gains and losses across multiple domains of HRQoL. (12,14) This is a valuable consideration in neurological disorders when the impact of a condition often traverses multiple domains. However, despite sharing a common purpose, utility instruments vary in many respects in content and structure; impacting the comparability of HRQoL assessments and utilities between instruments. (27) As a result, preference-based utility instruments often produce different evaluations of HRQoL and different utility estimates. (12,28,29) The magnitude of discordance in utilities has important implications for interpreting HRQoL status. The concern is that utility scores, and hence potential evaluations and decisions, may vary simply according to the choice of instrument. (30,31)

Despite a growing body of evidence regarding the performance of various utility instruments in the literature, head-to-head comparisons in persons with neurological conditions are limited. No prior study has evaluated the performance of utility instruments in relation to the common symptoms and functional limitations shared by neurological

conditions as a whole. Studies in general populations and other health conditions stress the importance of understanding how conceptual and methodological differences in preference-based instruments influence the evaluation of HRQoL. (27,28,32,33) Assessments of construct validity tend to support the idea that utility instruments measure HRQoL to some degree, (34-36) but there is insufficient information to guide the selection of instruments for the research task at hand. (37) As such, there is a need for evidence-based information regarding the potential applications and limitations of HRQoL measures in the presence of the health and life impacts associated with neurological conditions. (6)

This study compared the performance of two preference-based health related quality of life instruments: the Health Utilities Index - Mark 3 (HUI3) and the Short Form-6D (SF6D) in Canadians with neurological conditions. The study evaluated agreement/discordance between the SF6D and HUI3. It also assessed the sensitivity of the SF6D and HUI3 to HRQoL domains relevant to describing the impact of neurological conditions against a condition-specific HRQoL tool – Quality of Life in Neurological Disorders Measures (Neuro-QoL).

Structure of the Thesis:

This Thesis is in manuscript form. The format of the Thesis is as follows. 1) Introduction, 2) Background, 3 & 4) Formatted manuscripts, and 5) General conclusions. The background provides an overview of preference-based utility instruments and compares and contrasts the SF6D and HUI3 with respect to their methodological and evaluative qualities. The manuscripts replace the traditional methods and results chapters and address the research questions: 1) To what extent do the SF6D and HUI3 evaluate HRQoL domains relevant to individuals with neurological disease, disorders, and injuries, and 2) to what extent do evaluations of health related quality of life by the SF6D and HUI3, in persons with neurological conditions, agree? The general conclusions chapter outlines the implications for future research, in consideration of the limitations of the study.

CHAPTER 2. BACKGROUND

Preference-based health related quality of life (HRQoL) instruments (also commonly known as multi-attribute / utility / index instruments) are a unique class of generic HRQoL instruments designed to evaluate health in terms of “utility”. Utility instruments describe the HRQoL of an individual on a theoretical zero to one scale; describing the relative value of health in comparison to death (0.00) and perfect health (1.00). Scores less than zero are possible, and represent health states worse than dead. (12,22,27)

Utility instruments are distinct from psychometric measures of HRQoL in that they incorporate economic utility theory; borrowing the notion of “preference” in order to compare and aggregate the impact of health states on quality of life. “Utility” is not a measure of the presence or absence of illness; rather, it reflects the desirability or “preference” for the health states, enabling quantification of HRQoL as “better” or “worse”. Preferences can be obtained *directly* from the participant, an individual experienced in the health state, or *indirectly*, from a representative community sample. An important consideration in interpreting the utility scores of indirect-preference instruments is that the preference ascribed to a particular health state is not that of the individual under assessment – but rather the preference of a community sample. Community (indirect) preferences are deemed important for cost effectiveness analysis based on a utilitarian rationale that society's preferences should influence resource allocation decisions. (14,15,20 27,38-40)

2.1 PREFERENCE-BASED HRQoL INSTRUMENT USE

Preference-based utility instrument use is widespread in a broad spectrum of general disease categories. However, the use of utility instruments in neurological conditions represents only a small percentage of the studies to date. (27) The use of preference-based HRQoL measures has expanded over the past two decades because of their brevity, ease of use, and suitability in economic evaluations. (40) Although the development of utility instruments has its roots in economics, only 15% of these

instruments' historical use is in economic evaluations; the majority of their use is as generic HRQoL measures. (27)

The appeal of utility instruments is their ability to provide a single summary score across multiple health domains; a trait that lends itself well to assessments of HRQoL and the need to balance effects across physical, mental, and socio-emotional domains. In this way utility instruments are able to simultaneously weigh the net gain (or loss) in HRQoL by summarizing improvements in some domains and detriments in others. This is an important consideration in conditions where the impact on HRQoL spans multiple domains or in assessing interventions that may provide benefit in some health dimensions at the expense of others. Additionally, utility instruments are promoted as instruments that permit broad comparisons across populations and interventions. (23) These features have facilitated the ready inclusion of utility instruments in population surveys, research, and clinical applications across a broad spectrum of disease categories. (27)

However, utility instruments also have limitations. HRQoL assessment is constrained by the content of an instrument's health classification system. Even small differences in the content of HRQoL domains included in an instrument can alter the evaluation of HRQoL and utility estimates substantially, resulting in differences in agreement if applied to the same population. (20)

2.2 DIFFERENTIAL MEASUREMENT OF HRQoL

Preference-based instruments are comprised of two constituent parts: 1) a descriptive health classification system that defines and describes the health states of interest, and 2) a scoring system that assigns a value to the health states and converts the descriptive system into a summary score using a scoring algorithm. (22,27) Evaluations of HRQoL are influenced by interactions between an instrument's descriptive health classification system and the scoring system.

Even though utility instruments are designed to measure the same concept and have a similar purpose, they vary in almost every respect in regard to their descriptive systems and scoring systems. As a result, utility instruments, including the SF6D and HUI3, often produce very divergent evaluations of HRQoL. Table 2.1 compares and

contrasts the SF6D and HUI3 with respect to differences in the descriptive and scoring systems. *(See Appendices A and B for instrument details)*

2.3 DESCRIPTIVE HEALTH CLASSIFICATION SYSTEMS OF THE SF6D AND HUI3

Differences in the descriptive systems impact the way instruments define and describe health. The SF6D and HUI3 descriptive systems are based on different models of health. They also differ within their content and structure including: the type and number of health domains assessed, items per domain, and response levels included; as well as differences in the contextual framing of items.

2.3.1 Operationalization of HRQoL

Differing models and operationalization's of "health" by the SF6D and HUI3 influence the performance of each instrument in evaluating HRQoL in individuals with neurological conditions. (27) The HUI3 classification system is based on a narrower "within-the-skin" definition of health (32,41,42) that focuses on actual impairments; defining health status in terms of capacity (i.e. vision, speech, and ambulation) but omits social interaction considered outside the skin. (29) Whereas the SF6D, based on the SF-36v2, has a broader definition of health and, in addition to measuring physical and mental components of health, also measures emotional role and social functioning. (27) The implications of this are illustrated in the literature where the HUI3's within-the-skin model of health lends itself well to evaluating physical impairments and the SF6D tends to perform more optimally in outcomes that impact social functioning. (30,36)

2.3.2 Instrument content and structure

The SF6D and HUI3 instruments contain only a few domains defined in loosely similar ways and therefore measure different components of health with limited and imperfect overlap. (12,27) Only physical function, mental/emotional health and pain dimensions are "similar" in the SF6D and HUI3 – the remaining attributes are unique to their respective instruments. Even among the "similar" dimensions; the attributes included are often conceptually different. Within the broad

domain of mental and emotional health, the SF6D's mental health domain and HUI3s emotion domain are generally consider "similar". However, the HUI3 inquires about happiness and interest in life whereas the SF6D inquires about being downhearted and tense.

Omissions of important health domains or limited attribute item levels can also lead to variation in the range and severity of the potential health states captured and contribute to ceiling and floor effects. (43) For example, the ceiling effects often exhibited by the HUI3 are suggested to be due to the exclusion of important, "outside the skin" HRQoL domains such as vitality, role limitations, and ability to engage in vigorous physical activity; which would help discriminate more subtle impairments in healthier populations. The SF6D includes these domains and as a result it is able to better differentiate health states with a greater spread of values between utility scores of 0.90 and 1.00. (41)

2.3.3 Contextual framing differences

Differences in the contextual framing of questions also influences the characteristics of HRQoL domains captured in persons with disabilities. (44) This is especially apparent in elements of HRQoL that are more subjective in nature, such as items that elicit responses based on social functioning and effect on daily life rather than functional limitations alone. (39) In comparison to the HUI3, the SF6D permits individuals to gauge the impact of impairment; whereas the HUI3 items captures HRQoL states using statements that leave little room for subjective evaluation. (22) For example, the HUI3 captures information regarding ambulation with statements such as: (*"Able to walk around the neighbourhood with walking equipment, but without the help of another person"*) whereas respondents of the SF-36 are asked to indicate (*"Does your health now limit you in these activities? If so, how much? Yes, limited a lot; Yes, limited a little; No, not limited at all"*). (45) Framed in this way, an individual responding to the SF6D items has room to reflect on the impact of impairment on daily life, rather than simply indicating a functional limitation as in the HUI3.

Variations in the time frame reference between instruments will also affect how individuals differentiate between “usual” and “present” health status which can affect the stability and accuracy of capturing events; especially when considering differences between acute, episodic, or chronic conditions. (44) The SF-36v2, upon which the SF6D is based, frames questions either as “a typical day” or in the past four weeks whereas the HUI3 asks respondents to consider their “usual level of ability or disability”.

2.3.4 Comprehensiveness of the SF6D and HUI3 descriptive systems

The comprehensiveness of the SF6D and HUI3 will not only impact accurate evaluations of HRQoL it will also impact agreement between the instruments. A number of recent studies have explored the merits of particular HRQoL domains important to specific conditions, such as multiple sclerosis (MS), or neurological conditions as a whole. While the salience of these domains in neurological conditions differs somewhat between studies, common domain themes emerge and include: fatigue, social relationships and roles, occupational roles (including school, work, housework, and leisure), physical mobility (including ambulation and dexterity), balance, cognition, and mental and emotional health (including depression/anxiety and mood domains). (6-8,10,46) Neither the SF6D nor HUI3 contain all elements of this list. Each instrument does contain some domains to a greater or lesser extent. Both instruments are missing many elements of HRQoL considered relevant to neurological conditions such as balance and fatigue. As a result, their evaluations of HRQoL will not directly capture these issues, which will reduce the accuracy and validity of their respective utility scores and will affect agreement. (7)

2.4 SCORING SYSTEMS OF THE SF6D AND HUI3

Differences in each instrument’s scoring system means that valuations of health and aggregation of scores will differ between the instruments. The scoring system of a utility instrument uses economic methods to determine the value (or weighting) of health states. (43) The scoring system converts the descriptive system into a summary utility score and can vary by preference elicitation technique, method of preference

measurement, and the scoring function (algorithm) applied. Three common valuation techniques are: standard gamble (SG), time trade-off (TTO), and visual analogue score (VAS); each method yields different values. Preference valuation studies differ in terms of respondents, the size of the sample, method used to administer the questionnaire, and complexity of health states valued. (19) Both the SF6D and HUI3 use indirect preference weights obtained from valuation studies using community samples: the HUI3 from a Hamilton, Ontario general population sample (n=504) (22), and the SF6D from a UK general population sample (n=836). (26,32,47) The SF6D and HUI3 also differ in their scoring function and in the complexity of the health states that are used to derive the weights. The HUI3 scoring function was developed by rating single-deficit states using VAS and standard gamble techniques in a multiplicative model. In contrast, weights for the SF6D were derived from standard gamble valuations that included complex, multi-deficit states but used a linear additive model to aggregate domain scores. (23,26,32)

An instrument's scoring function determines the range of utility scores. The utility range between the SF6D and HUI3 differ dramatically. Although both measures have a theoretical utility range of 0.00 to 1.00, the range of the SF-6D is only 0.30 to 1.00; whereas the HUI3 can describe ill health into a range of - 0.36. The narrower range of the SF6D contributes to floor effects and an inability to detect deterioration in lower health states; often considered a limitation of the SF6D. Conversely, the HUI3 is suggested to better describe more severe burden of illness with fewer floor effects, particularly in the worse-than-dead range. (13,22,33,41,48)

2.5 INSTRUMENT DIFFERENCES AND IMPLICATIONS FOR HRQoL MEASUREMENT

Different utility instruments often produce different utility estimates for the same individuals. Some of the variation in utility estimates is random, a small amount is attributable to preference elicitation methods and different scoring systems; however, by and large the majority of the discordance is attributable to differences in the descriptive health classifications systems. (27) Differences in an instrument's descriptive health classification system mean that respondents are not prompted or able to report their health status in a similar way between instruments. Different instruments will collect information on different dimensions of health, in a different order, and with different levels of detail

and opportunity to reflect on impact. As a result, different health characteristics will be included, emphasized, or omitted resulting in different evaluations of HRQoL and utility scores estimates. Not surprisingly, applying different utility instruments to the same individual can result in very different utility scores being generated for the same HRQoL state. (13,27)

Inadequacy in a descriptive health classification system, in relation to a particular population, may limit or restrict an instrument's ability to measure key attributes of health (validity), to distinguish important differences between health states (sensitivity), and to measure important change (responsiveness). (17,27) This leads to questioning of the overall interpretability and meaningfulness of scores generated by utility instruments. (8) The effect of the different descriptive systems on overall utility assessment and validity of utility scores in various populations remains a critical and unresolved issue. (27)

2.6 AGREEMENT BETWEEN THE SF6D AND HUI3

Despite the research showing reasonable associations between the utility scores of different instruments, the extent of *agreement* between instruments is much poorer in comparison. (49) Of primary concern is the extreme range of utility obtained in the same individual (within-subject) by different instruments. Evidence of differential HRQoL measurement and lack of agreement between utility instruments is common, irrespective of the instrument under review and the population studied. Assessments of construct validity tend to support the idea that utility instruments measure health related quality of life to some degree. However, significant systematic differences exist in the distribution of scores and evaluation of HRQoL domains. (27-29,34,36,44,50-53)

The collection of evidence from comparative studies between utility instruments shows that: 1) different instruments provide different utility estimates in the same population, 2) utilities provided by different instruments are not equivalent, 3) instruments differ in their sensitivity to different health attributes and/or severity of impairment 4) instruments are not equally responsive to change, 5) utilities generated by different instruments may have considerable effects on economic evaluation studies, and 6) and comparisons based on utilities from different instruments warrant caution. (12,27,29,33,43,53)

Mean differences, intra-class correlations (ICC), and Bland-Altman plots are the most common analysis of agreement / discordance between utility measures. (*Details of methods to assess agreement between instruments and the limitations of the methods are discussed in detail in Appendix D*). The ICC estimates the proportion of *between* subject variations in relation to total variation where 1 represents perfect agreement and 0 - no agreement at all. (54) The ICCs between the SF6D and HUI3 reported in the literature range from 0.28 to 0.44, indicating only poor to fair agreement. (32,41,50,55,56) Bland-Altman plots graphically represent the measurement bias between two measures. In the literature, regardless of the instruments compared, these plots typically show that discordance varies systematically through the range of utility and have wide limits of agreement. (44,57-60) A few studies investigated the possibility of translations between instruments so that scores from one measure can be used to estimate the scores in another, but conclude that translations or linear “crosswalks” between instruments result in “low precision”. (28,32,51)

2.7 FACTORS AFFECTING UTILITY DIFFERENCES BETWEEN THE SF6D AND HUI3

A small number of studies have ventured to describe participant characteristics associated with utility difference. (61-63) These factors can be broadly organized into three categories: 1) level of HRQoL status and severity of impairment 2) socio-demographic characteristics, and 3) condition characteristics.

1. HRQoL status and severity of impairment

The SF6D and HUI3 are not equal in their ability to describe HRQoL through a full range of disutility. Consistent with the differing score ranges and frequently cited ceiling and floor effects, the degree of discordance between instruments is associated with level of utility and is generally more extreme at the upper and lower range of utility scores. (32,57,58)

2. Socio-demographics

HRQoL status varies by factors such as age, sex, race, education and income. Not only are these factors determinants of health-related quality of life; utility measures are differentially influenced by these factors. While the SF6D is

found to be more sensitive to gender than the HUI3, the HUI3 is more influenced by education and sensitive to socio-economic disparities than the SF6D. (61,64-66)

3. Condition characteristics

How a condition is acquired, whether it is stable, degenerative, or episodic or whether the impairments are of a more physical, mental, or social nature will have a unique impact on a person's health related quality of life, and subsequently on the evaluation of HRQoL due to the different descriptive systems of utility instruments. For example, in patients affected by a stroke, Pickard and colleagues (2005) found that changes in utility scores in the SF6D were more strongly associated with changes in mental and social aspects of health, whereas the HUI3 was more strongly related to changes in measures of disability and activities of daily living. (30) This is consistent with the "outside-the-skin" / "within-the-skin" emphasis of each instrument respectively.

2.8 IMPLICATIONS OF MEASUREMENT DISCORDANCE

Studies have found substantive discrepancies in quality adjusted life years (QALY's) based on utilities from the SF6D and HUI3. (5,30,31,37) Because the HUI3 has a larger range of scores, it tends to appear more responsive to change and generally produces larger change scores than the SF6D. Larger change scores tend to lead to more favourable QALYs. For example, Marra and colleagues (2007) modeled the effects in incremental QALYs generated from different utility instruments in a sample of Rheumatoid Arthritis patients. It was found that a particular treatment was favoured in 91% of their simulations using the HUI3; whereas the same treatment was only favoured in 12% of simulations using the SF6D. (31) The inference is that any substantive discordance in utilities could have potential economic and treatment implications. (18,30,31,37) For example, Vossius and colleagues (2009) evaluated the effect of a drug treatment on Parkinson's disease using the SF6D and determined that the drug resulted in increased health status but that the incremental cost-effectiveness ratio was high (unfavourable). (5) Based on the evidence this raises an important question: if the HUI3 had been used instead of the SF6D to generate QALYs would the cost-effectiveness have been more acceptable?

Between-instrument discrepancies can affect HRQoL evaluations and comparisons of benefit, and have the potential to unduly influence clinical decision-making; policy recommendations; and funding decisions. (14,16) Quite simply, interventions (or populations) could be favoured or disadvantaged because a particular utility instrument captures more or less of a health domain specific to the intervention (or population), leading to systematic bias in economic evaluations. (43) Given the substantial between-instrument differences between the SF6D and HUI3, these examples highlight the need for an in-depth exploration of how these instruments perform in individuals with neurological conditions.

2.9 LIMITATIONS IN THE LITERATURE

A review of the current evidence highlights gaps in the research in evaluating the performance of preference-based utility instruments in persons with neurological conditions. While previous research has led to an understanding of conceptual and methodological differences in preference-based instruments and their collective influence on the evaluation of HRQoL, few studies have compared the performance of the SF6D and HUI3 in evaluating HRQoL in neurological conditions. Only three comparative agreement/discordance studies that included both instruments were found. Picard and colleagues compared the responsiveness of the EQ-5D, HUI2, HUI3, and SF6D in individuals with Stroke. (30) Fisk and colleagues (2005) compared the feasibility and psychometric properties of the HUI3, EQ-5D and SF6D in individuals with MS. (34) Langfitt and colleagues (2006) compared responsiveness and validity of the EQ-5D, HUI2, HUI3, and SF6D in individuals with chronic epilepsy. (36) Each study highlights difference between the instruments with respect to strengths of association, discriminative ability between levels of disability, and differing responsiveness. Lacking in the current literature, with respect to the SF6D and HUI3, is:

1. A quantification of between-instrument discordance across a wide range of neurological conditions, symptoms and impairments
2. An evaluation of how well each instrument captures the physiological, psychosocial, and emotional consequences of neurological conditions

3. A comprehensive description of the factors that contribute to differential HRQoL measurement in individuals with neurological conditions.

There is a need for research to describe and evaluate the performance of preference-based utility instruments in neurological conditions; not only to inform the choice of measure and improve interpretation of their findings, but also to inform decision-making so that individuals are not disadvantaged by the choice of instrument.

2.10 SUMMARY

Differences exist in the way preference-based health related quality of life instruments define, describe, and value HRQoL. Differences in the instruments descriptive systems and scoring systems impact evaluations of HRQoL. As a result, respondent and condition characteristics interact with the differing conceptual and methodological qualities of the instruments to produce different evaluations of HRQoL for the same person. What is not well understood are a number of related issues: how well the SF6D and HUI3 capture HRQoL domains relevant to evaluating HRQoL in individuals with neurological conditions; the consequences of differential HRQoL measurement; what factors impact discordance; and under what circumstances the choice of measure may favour or disadvantage particular populations or interventions?

2.11 PURPOSE OF THE STUDY

The purpose of this study is to provide a comprehensive analysis of the performance of the SF6D and HUI3 in individuals with neurological disease, disorders and injuries. This includes evaluating if the instruments capture HRQoL domains relevant to neurological conditions. A significant component of appraising performance is assessing the level of agreement between instruments. Comparing the patterns of concordance/discordance between the SF6D and HUI3 will provide information regarding strengths and gaps in each instrument's assessment of HRQoL in this population. Additionally, understanding what factors affect utility differences will help identify the potential measurement bias each instrument may impose.

2.12 OBJECTIVES

The objectives of this study were to:

1. Assess the degree to which the SF6D and HUI3 were sensitive to variation in domains of HRQoL of life relevant to persons with neurological conditions
2. Describe discordance between the SF6D and HUI3 overall and how it varied by participant characteristics, type of neurological condition, and HRQoL domains relevant to persons with neurological conditions.

Each objective is addressed in a separate chapter (manuscript). Chapter 3. “Sensitivity in Measuring Health Related Quality of Life in Canadians with Neurological Conditions”, addresses Objective 1. Chapter 4. “Discordance in Utility Measurement in Canadians with Neurological Conditions: A Comparison of the SF6D and HUI3”, addresses Objective 2.

Table 2.1: Comparison of the SF6D & HUI3 by descriptive health classification system and scoring system characteristics

	SF6D	HUI3
DESCRIPTIVE SYSTEM		
<i>Definition of Health</i>	Based on WHO definition of health that includes physical, mental and social functioning.	Based on a “within-the-skin” definition of health (omits social interaction)
<i>Health Domains</i>	6 attributes of health status: physical functioning, role participation (combined role-physical and role-emotional), social functioning, bodily pain, mental health, and vitality	8 attributes of health status: vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain
<i>Items per domain</i>	1-3 items	1 item
<i>Response levels</i>	4-6 response levels	5-6 response levels
Contextual differences		
<i>Recall Period</i>	Some questions: ‘typical day’ Some questions: ‘in past 4 weeks’	'Usual health' used in LINC study
<i>Subjective /objective framing of an item</i>	Designed to capture “perceived” health states and impact by using subjective wording such as: (i.e. <i>Does your health now limit you in these activities? If so, how much? Yes, limited a lot; Yes, limited a little; No, not limited at all</i>)	Designed to solicit health states objectively such as: <i>“Able to walk around the neighbourhood without difficulty, and without walking equipment”</i>
SCORING SYSTEM		
<i>Preference measurement & sample</i>	Indirect community preference weights were obtained from a sample of the UK general population (n=836) who were asked to value a sample of 249 possible health states	Indirect community preference weights were obtained from a random sample of the general population in Hamilton, Ontario (n=504)
<i>Preference elicitation</i>	Interviewer-administered rating of multiple-deficit health states using Standard gamble in a two-stage cascade technique	Interview-administered rating of single deficit health states using visual analogue scale (VAS); scores converted to standard gamble valuation using a statistical power transformation
<i>Scoring function</i>	A linear additive regression model was fit by ordinary least squares with SF-6D item levels and interactions as covariates.	Multiplicative utility model
RANGE AND VARIANCE		
<i>No. of health states</i>	18,000 unique health states	972,000 unique health states
<i>Range of score</i>	Utility score: 0.30 to 1.00	Utility score: - 0.36 to 1.00
<i>Presence of Ceiling and Floor Effects</i>	Common floor effects in populations with more morbidity, but fewer ceiling effects.	Ceiling effects in healthier populations but describe wider range in lower health states.

CHAPTER 3

(Manuscript)

Sensitivity in Measuring Health-Related Quality of Life in Canadians
with Neurological Conditions:

A Comparison of the SF6D and HUI3

ABSTRACT

BACKGROUND

Health related quality of life (HRQoL) is an important outcome in persons with neurological conditions, as it contributes to a greater understanding of disease consequences and treatment effects. Differences in the way preference-based HRQoL (utility) instruments define, describe and value HRQoL affect the validity, interpretability, comparability, meaningfulness, and agreement of utilities generated by these instruments. Although the performance of these instruments has been studied before, they have not been evaluated in a diverse sample of neurological conditions.

OBJECTIVES

The study assessed to what extent the SF6D and HUI3 are sensitive to variation in domains of HRQoL relevant to persons with neurological conditions. Specifically, the instruments were compared to the Neuro-QoL measures to assess i) the ability of the SF6D and HUI3 to differentiate HRQoL domains relevant to persons with neurological conditions and ii) the ability of the SF6D and HUI3 to discriminate among degrees (severity) of HRQoL impairment.

METHODS

This study involved the analysis of cross-sectional population survey data collected as part of the LINC Study (Living with the Impact of a Neurological Condition). Self-reported data were collected on the burden and impact of neurological conditions on participants' everyday lives. Associations between the SF6D and HUI3 HRQoL domains relevant to persons with neurological conditions (Neuro-QoL) were evaluated using Spearman correlation matrices. The ability of the SF6D and HUI3 to discriminate between degrees of HRQoL was assessed using Analysis of Variance (ANOVA).

RESULTS

The data of 776 individuals with scores for both the SF6D and HUI3 were analyzed. HRQoL of the sample is low with mean utility scores of HUI3: 0.47 (95% CI 0.45, 0.49) and SF6D: 0.62 (95% CI 0.62, 0.63). The SF6D and HUI3 were both sensitive to a variety of HRQoL domains relevant to persons with neurological conditions and were able to distinguish between degrees of HRQoL impairment. However, the SF6D and HUI3 measure a different array of HRQoL domains, with differing sensitivity and discrimination.

CONCLUSIONS

Utility instruments are useful in providing a summary score of HRQoL that weighs the gains and losses in HRQoL across multiple domains simultaneously. The SF6D and HUI3 provide unique information on the impact and burden of neurological conditions on the everyday lives of Canadians.

3.1 INTRODUCTION

Neurological disease, disorders, and injuries are a significant cause of disability and disease burden in Canada. (1,2) Very few neurological conditions are curable and often the impact of neurological conditions extends beyond physical and cognitive impairments and affects social relationships and roles with family, friends, and the wider community. (6) As a result, the treatment emphasis is often on managing the impact of symptoms and the maintenance of health-related quality of life. (3-6,8,10,67)

Health-related quality of life (HRQoL) is an important outcome to capture the “impact” of disabling conditions that are life-long and multidimensional. (7,8,68) Assessment of HRQoL contributes to a greater understanding of disease consequences and treatment effects on everyday life. (6) A variety of generic, condition-specific, and preference-based utility instruments are used to evaluate HRQoL in persons with neurological conditions; each instrument having its own strengths, limitations, and benefits. (6)

Health-related quality of life instruments can be broadly categorized into psychometric instruments (generic and condition-specific instruments) and preference-based utility instruments. Generic, psychometrically based instruments, such as the widely used Rand Corporation’s Short-Form 36 (SF-36v2), are useful in describing a number of HRQoL domains considered broadly relevant to most individuals; but are criticized for a lack of sensitivity in capturing specific issues relevant to specific conditions. (12,69) Condition-specific instruments, such as Stroke Impact Scale (SIS) (ref) (70); Multiple Sclerosis Impact Scale-29 (MSIS-29) (71); and the Quality of Life in Epilepsy-31 (QOLIE-31) (72) are designed to provide a detailed picture of disease impacts that are relevant to the condition in question. As such, they tend to be more sensitive to issues, symptoms, functioning and degrees of condition severity, and are able to detect smaller changes in HRQoL. However, disease specific measures are limited in that they do not provide an overall picture of the whole person; they may lack sensitivity to co-morbid conditions; and limit cross-condition comparisons. (8,12,68) A major limitation of psychometrically based (generic or condition specific) instruments is their inability to provide an overall summary score that weighs gains and losses in multiple HRQoL domains simultaneously. (12)

Preference-based health related quality of life instruments (also commonly known as multi-attribute / utility / index instruments) are a unique class of standardized HRQoL instruments that evaluate HRQoL in terms of utility. The Short-Form 6D (SF6D) (26) and the Health Utilities Index – Mark 3 (HUI3) (22,23) are examples of these instruments. Utility instruments provide a summary score of HRQoL that is interpreted on a theoretical zero to one scale; where 1.00 represent perfect health, and 0.00 - death. Utility instruments are distinct from psychometric measures, in that they incorporate preferences that reflect the desirability of a health state as “better” or “worse”. (12-15)

The use of utility instruments has expanded over the past two decades. This has been facilitated by their brevity, ease of use, suitability for economic evaluations and in response to a need to compare health across populations and treatments. (27,40,73) Utility instruments have been included in population surveys and clinical applications across a broad spectrum of disease categories; however, their use in neurological conditions represents only a small percentage of the studies to date. (30,34,36,74-80)

There are pros and cons to all measurement approaches. Limitations and differences exist between utility instruments that can impact the accuracy and validity of evaluation of HRQoL in neurological conditions. These include variation in the comprehensiveness and inclusion of different HRQoL domains that are more or less relevant to neurological conditions, variation in sensitivity and discrimination between levels of neurological disability, and differing responsiveness to change. (7,8,27,30,34,36) Research in general populations and other conditions stress the importance of understanding how methodological and content differences in utility instruments affect the evaluation of HRQoL in specific diseases and conditions. Currently there is a lack of evidence to guide the choice of utility instrument for use in persons with neurological conditions. (27)

The intent of this study was to assess the degree to which the SF6D and HUI3 are sensitive to variation in domains of HRQoL relevant to persons with neurological conditions. No prior study has evaluated the cross-sectional sensitivity of the SF6D and HUI3 to the common symptoms and functional limitations shared by neurological conditions as a whole. Meeting this objective will guide answers to questions such as: Do the SF6D and HUI3 capture domains important in describing HRQoL in persons with

neurological conditions? Can the SF6D and HUI3 discriminate between different degrees of HRQoL impairment and symptom severity? How comprehensive are the SF6D and HUI3 instruments for assessment of HRQoL in persons with neurological conditions?

3.2 METHODS

This study assessed the sensitivity of the SF6D and HUI3 to variation in HRQoL domains relevant to persons with neurological conditions. It involved the analysis of data from a sample of Canadians with neurological conditions drawn from the cross-sectional population survey collected as part of the LINC Study (Living with the Impact of a Neurological Condition) and includes data that describes the health burden and impact on participation in everyday life of people with neurological conditions. (81) The study compared the sensitivity of two preference-based utility instruments: the Short-Form 6D (SF6D) (26) and the Health Utilities Index – Mark 3 (HUI3) (22,23) to the recently developed Quality of Life in Neurological Disorders (Neuro-QoL) measures. (10,11) The Neuro-QoL is designed to evaluate multiple HRQoL concerns common to, and applicable for, many neurological conditions. For the purpose of this study, the domains included in Neuro-QoL were used to represent HRQoL domains relevant and important to the evaluation of HRQoL in persons with neurological conditions. (67) The study data was analyzed using STATA 12 (StataCorp). (82)

3.2.1 Sample and data collection

A sample of 948 Canadians 17 years of age or older, with at least one identified neurological condition, participated in the cross-sectional population survey administered as part of the LINC Study (Living with the Impact of a Neurological Condition). The sample was a non-random (convenience) sample and was recruited by the Neurological Health Charities Canada (NHCC) and NHCC member organizations. Recruitment strategies included a variety of word-of-mouth invitations distributed through relevant organizations, known databases, and registries of people with neurological conditions. (81)

Participants completed a comprehensive survey on the burden and impact of neurological conditions on everyday life between September 14, 2011 and July 1, 2012.

The majority of respondents (89%) completed the survey online using Opinio, (83) a web-based survey application. The remaining respondents completed the survey on paper (8%) or by telephone (3%). Self-reported data on diagnosis and socio-demographics such as age, sex, marital status, employment status, and education were collected. Clinical information included time since diagnosis, self-rated health, indication of a decline in health in past year, and presence of other co-morbid conditions (diabetes, heart disease, and depression). Data was exported from Opinio, examined and cleaned using Excel (84), and derived variables were calculated using SAS (85) for standard measures, scales and indices. Missing data was handled as per each instrument's standard protocol.

Data of the 776 individuals who had utility scores for both the SF6D and HUI3 was analyzed. From the original sample of 948, 172 individuals were excluded using pairwise deletions due to 95 missing SF6D and/or 134 missing HUI3 scores. There were some differences between survey completers and non-completers; those with missing HUI3 and/or SF6D utility scores were more likely to have an early-onset/traumatic condition (degenerative conditions were associated with missing HUI3 values only). In addition, individuals missing HUI3 scores were more likely to have greater anxiety, lower self-efficacy, and were more likely to need help with self-care.

3.2.2 Measures of HRQoL

The LINC population survey was constructed using a series of validated scales and standardized instruments selected to collect data about the impact of neurological conditions. The analysis used data from the Short-Form 6D (SF6D), (26) Health Utilities Index – Mark 3 (HUI3), (22,23) and Quality of Life in Neurological Disorders (Neuro-QoL) measures. (10,11)

The SF6D and HUI3 are generic, preference-based multi-attribute health state classification and utility scoring systems. Each instrument can provide a profile of health describing the extent of disability by HRQoL domain and also summarize HRQoL as a single utility-based index score. (22,26) Neuro-QoL is a set of HRQoL measures applicable for people with common neurological conditions. (67) Neuro-QoL was developed as a clinically relevant and psychometrically robust QOL assessment tool for neurologic disorders to enable clinical researchers to compare the QOL impact of different

interventions within and across various neurological conditions. (10,11)

Health Utilities Index – Mark 3(HUI3) (www.healthutilities.com) (22,23)

The HUI3 describes 8 domains of HRQoL: vision, hearing, speech, dexterity, ambulation, cognition, emotion, and pain. Each domain has 5-6 levels of function. The instrument describes 972,000 unique health states. Health state valuation used standard gamble from a sample of 504 Canadians. HUI3 utility scores can range from -0.36 to 1.00. The HUI3 utility score is calculated using the multi-attribute scoring function and Canadian sample preference weights provided by Health Utilities Incorporated. The HUI3 has widespread use in clinical and general applications, has been included in major population health surveys in Canada since 1990, and is one of the most frequently used tools in economic analysis. (18,22) *(Further details in Appendix A)*

Short-Form-6D (SF6D) (derived from SF-36v2) (www.SF36.org) (26)

The SF6D describes 6 domains of HRQoL: physical functioning, role limitations (physical and emotional), bodily pain, vitality, social functioning, and mental health. Each domain has 1-3 items, which have 4 to 6 levels of function. The instrument describes 18,000 unique health states. Health state valuation used standard gamble from a sample of 836 individuals in the UK. SF-6D utility scores can range from 0.30 to 1.00. The SF6D utility score and health domains are calculated from 11 of the 36 items of the SF-36v2 completed in the survey, based on the SF6D preference-based algorithm and UK preference weights provided to the author by Brazier and colleagues. (26,86-88) The SF6D is more widely used in Europe and, unlike other instruments that are conceived and developed specifically as preference-based tools, the SF6D is unique in that its utilities are derived using an algorithm developed by Braizer and colleagues from 11 of the 36 items of the well established psychometric instrument – the Short Form-36 (SF-36v2). (19,26,27,45) Given that the SF-36v2 is the most widely used psychometric measure of HRQoL, the development of the algorithm to convert existing and future SF-36 results into utility scores has been seen as a valuable addition to utility measurement. (12,35) *(Further details in Appendix B)*

Quality of Life in Neurological Disorders measures (Neuro-QoL) (www.neuroqol.org)
(10,11)

Neuro-QoL is a condition-specific outcome measure designed to generate clinically relevant and psychometrically robust health-related quality of life (HRQL) assessment data across a broad spectrum of symptoms and issues relevant to neurological conditions. It was developed in response to a lack of specificity in generic instruments; a lack of a common condition-specific instrument used in neurological conditions; and a desire to better capture a person's experience of neurological disease symptoms, treatment side effects, functioning and well-being. (10,11) While Neuro-QoL is labeled as a health related quality of life measure, it does not provide a summary measure of HRQoL. Rather, Neuro-QoL contains a series of separate condition-specific psychometric scales of HRQoL domains relevant to many neurological disorders. Thirteen of the Neuro-QoL short form scales were included in the LINC study: upper extremity function, lower extremity function, cognition (general and executive function), depression, anxiety, fatigue, sleep disturbance, emotional and behavioral dyscontrol, positive affect and wellbeing, ability to participate in social roles and activities, stigma and satisfaction with social roles and activities. Each scale has 8-9 items describing the domain. Raw scores for each scale were converted to T-scores based on the scoring tables in the Neuro-QoL user manual. T-scores have a mean of 50 and standard deviation of 10. The scores are either standardized in reference to the US general population or a clinical population. Higher scores in a Neuro-QoL scale indicate that more of a domain is measured. For example, a high fatigue score indicates worse HRQoL as it relates to fatigue than a lower score, whereas, a high "lower extremity" score indicates better HRQoL in this domain. A 0.5 SD change in T-score is considered a clinically relevant difference. The scales were developed using item-response theory. As a new measure, validity studies are limited – but it was recently validated in adults with epilepsy and showed good evidence for internal consistency, test–retest reliability, convergent validity, and responsiveness to change over several months. (46,89) (*Further details in Appendix C*)

HRQoL domains measured by the SF6D, HUI3, and Neuro-QoL

SF6D Domains (No. of domain levels ^a)	HUI3 Domains (No. of domain levels ^a)	Neuro-QoL Scales ^b (Reference population ^c)
Physical functioning (6 levels)	Ambulation (6 levels)	Upper Extremity (GPR)
Role limitation (4 levels)	Dexterity (6 levels)	Lower extremity (GPR)
Social functioning (5 levels)	Cognition (6 levels)	Cognition-general concerns (GPR)
Mental health (5 levels)	Emotion (5 levels)	Cognition- executive function (GPR)
Pain (6 levels)	Pain (5 levels)	Positive affect and wellbeing (GPR)
Vitality (5 levels)	Vision (6 levels)	Ability to participate in social roles (GPR)
	Speech (5 levels)	Satisfaction with social roles (GPR)
	Hearing (6 levels)	Fatigue (NCS)
		Sleep disturbance (NCS)
		Depression (NCS)
		Anxiety (NCS)
		Emotional behavioural dyscontrol (NCS)
		Stigma (NCS)
<p>a In both instruments, domain levels are based on categories of functioning (Likert type scale). Domain level 1 represents full functioning in the domain. Functioning <i>declines</i> with increasing domain levels, where levels 4 to 6 represent severe HRQoL impairment.</p>		<p>b With the exception of the positive affect and wellbeing scale that has 9 items – all Neuro-QoL scales have 8 items per scale.</p> <p>c Scale references a general population (GPR) or scale references a neurological clinical sample (NCS)</p>

3.2.3 Statistical Analysis

Descriptive statistics were used to characterize the sample in terms of participant characteristics and HRQoL. Means and 95% confidence intervals (95% CI) were reported for continuous variables and frequencies and proportions (%) for categorical variables. Mean summary utility scores, standard deviations (SD), and 95% CI were reported for the SF6D and HUI3. Individual domain scores were calculated for the SF6D using preference weighted domain scores and for the HUI3 using the multi-attribute domain scores. Mean T-scores and SD were reported for the 13 Neuro-QoL scales.

To meet the study’s objective, and assess the degree to which the SF6D and HUI3 were sensitive to variation in domains of health related quality of life relevant to persons with neurologic conditions, the instruments were compared with respect to: i) the ability to differentiate HRQoL domains relevant to persons with neurological conditions and, ii) the ability to discriminate among degrees (severity) of HRQoL impairment.

Intercorrelations (Spearman’s Rho) between the utility instruments (SF6D and HUI3) and Neuro-QoL were compared to assess the degree to which the SF6D and HUI3 were sensitive to domains relevant to persons with neurologic conditions. Two comparisons were made: i) SF6D and HUI3 utility scores were compared to Neuro-QoL

domain T-scores to assess the overall relationship between the utility scores and Neuro-QoL domains and, ii) the relationship between individual SF6D and HUI3 domains (using the SF6D preference-weighted domain scores and HUI3 multi-attribute domain scores) and Neuro-QoL domains were assessed to gain an understanding of which utility domains were important in capturing the Neuro-QoL domains. Guyatt and colleagues' (1993) suggested correlation classifications, were used to interpret the magnitude of correlations: correlations greater than 0.50 represent a strong relationship; correlations between 0.35 and 0.50 represent a moderate relationship; and below 0.35 a marginal relationship between domains. (14) It was expected that there would be weaker associations between domains of dissimilar constructs and stronger correlations between domains of similar constructs.

The ability of the SF6D and HUI3 to detect different degrees of HRQoL impairment was assessed using analysis of variance (ANOVA). This is based on the assumption that individuals with different degrees of HRQoL impairment will have different utility scores, and that more impaired subjects would report lower HRQoL and vice versa. To test if the SF6D and HUI3 utility scores were able to reflect variation in HRQoL impairment, participants were stratified within each Neuro-QoL domain by “normal”, “mild” and “moderate to severe” impairment. ANVOA was used to test if the mean utility scores were statistically different between severity categories (F-test p-value <0.05).

3.3 RESULTS

Characteristics of the sample

The data of 776 individuals who had scores for both the SF6D and HUI3 were analyzed. Characteristics of the sample are presented in Table 3.1. The sample had a mean age of 50.8 years (95% CI 49.7, 51.9), was predominantly female (64.5%), was highly educated, and a high proportion of the sample (64%) did not work due to health or other reasons. Seventeen neurological diagnoses were represented in the sample. The average time since diagnosis was 17 years. Participants reported on average 1.5 (SD 0.83) neurological conditions per person.

HRQoL of the sample

The impact of neurological conditions on HRQoL was high, evidenced by low utility scores. The HUI3 produced a lower mean utility of 0.47 (95% CI 0.45, 0.49); compared to the SF6D mean utility of 0.62 (95% CI 0.62, 0.63). The mean T-scores and standard deviations for the 13 Neuro-QoL scales are presented in Table 3.2. The sample reported lower physical functioning (upper & lower extremity), greater difficulties in applied cognition (general concerns and executive function), less ability to participate in social roles and activities, and less satisfaction with social roles and activities compared to the reference general population. The domains of fatigue, sleep disturbance, emotional-behavioural dyscontrol, depression, anxiety, and stigma are comparable to the clinical neurological reference population, and only varied from the mean reference T-score by one or two points.

Ability to differentiate HRQoL domains relevant to persons with neurologic conditions

Spearman correlations between the SF6D and HUI3 and Neuro-QoL are shown in Tables 3.3 & 3.4. The degree to which the SF6D and HUI3 were sensitive to domains of HRQoL that are relevant to persons with neurologic conditions were compared by assessing the intercorrelations between the utility instruments and Neuro-QoL using Spearman correlation matrices. Utility scores of both instruments were moderately to strongly associated with most Neuro-QoL domains. The SF6D's summary utility score was more strongly correlated with Neuro-QoL domains than the HUI3 utility score; with a higher strength of associations with 9 of the 13 scales. With respect to the SF6D's domains – the social functioning and mental functioning domains were moderately and strongly associated with the majority of Neuro-QoL domains. The HUI3's emotion and cognition domains were also moderately to strongly associated with the majority of the Neuro-QoL domains.

Ability to discriminate between degrees (severity) of HRQoL impairment

Mean utility scores by normal, mild, and moderate to severe HRQoL impairment by Neuro-QoL domain are reported in Table 3.5. Higher levels of HRQoL impairment had lower utility scores for all Neuro-QoL domains. The SF6D and HUI3 were able to

discriminate between differing severities of impairment (normal, mild, and moderate-severe impairment) in all domains and between all severity levels (p-value < 0.001). However, the HUI3 showed greater magnitude of difference in utility scores between adjacent levels of Neuro-QoL domain impairment. For example the difference in utility between mild fatigue and moderate-severe fatigue was 17 points in the HUI3 but only 7 points in the SF6D.

3.4 DISCUSSION

This study is the first comparison of the SF6D, HUI3, and Neuro-QoL measures undertaken for a variety of neurological conditions. A strength of this study is the depth of the data that included three standardized HRQoL instruments. Based on the recently developed Neuro-QoL Measures, the SF6D and HUI3 were found to be sensitive to a variety of HRQoL domains relevant to neurological conditions and both instruments were able to distinguish between degrees (severities) of HRQoL impairment. However, the SF6D and HUI3 did not measure HRQoL in the same way. The SF6D and HUI3 differed in mean utility scores, in the domains they each preferentially captured, and differed in the magnitude of utility difference between degrees of impairment.

The sensitivity of an instrument relates to its discriminatory power to detect different elements and different levels of HRQoL status at a given point in time. (49) The SF6D and HUI3 were both sensitive to domains relevant to neurological conditions (Tables 3.3 and 3.4) However, each instrument contains domains that were more or less sensitive in describing the impact of neurological conditions. Based on the pattern of correlations, the HUI3 had greater association to outcomes of physical and cognitive functioning, whereas the SF6D had greater association with outcomes that impact social functioning and are more subjective in nature such as stigma and fatigue. This finding is similar to other studies of neurological conditions, where the SF6D and HUI3 show strengths and limitations in domains consistent with the differing “within-the-skin” – “outside-the-skin” domain emphasis of each instrument. (30,34,36)

The SF6D and HUI3 were both able to discriminate between degrees of HRQoL impairment. However, their sensitivity was not equal. The HUI3 showed greater magnitude of difference in utility scores between degrees of HRQoL impairment stratified

by ‘normal’, ‘mild’, and ‘moderate to severe’ impairment. (Table 3.5) These differences suggest that the HUI3 may capture more subtle changes in HRQoL than the SF6D. However, this benefit may not be consistent throughout the range of utility. Despite the narrower range of differences, the SF6D *was* sensitive to describing the nature and severity of HRQoL impairment in the sample. Further, the SF6D had stronger associations between several neurologically relevant HRQoL domains absent from the HUI3’s descriptive system (e.g. vitality, social functioning, role limitations), which enabled the SF6D to capture nuances of neurological impairment that the HUI3 could not. In neurological conditions, where the impact of a condition extends beyond functional limitations and impacts social aspect of everyday life, the SF6D’s consideration of “outside-the-skin” domains is important to describing the HRQoL. The relevance of these domains is seen in the proportion of individuals in the lowest utility range, who reported the poorest functioning in SF6D domains of role participation, vitality, and social functioning (88%, 57%, 41% respectively) rather than in physical functioning (34%).

An appeal of utility instruments is in their ability to provide a single summary score across multiple health domains; a trait that lends itself well to assessments of HRQoL and the need to balance effects across domains. This is important in conditions where the impact on HRQoL spans multiple domains or in assessing interventions that may provide benefit in some domains at the expense of others. For example, Neuro-QoL provides superior detail regarding the 13 individual HRQoL domains; however it is impossible to discern the overall impact of these physical, emotional, cognitive, and social impairments at the individual or sample level. Utility instruments are useful because they can sum the impact on individual domains and, in theory they permit broad comparisons across populations and interventions because of their index (ratio) scale. (23) Comparison of outcomes across studies and populations is important when trying to determine the comparative level of HRQoL or the comparative effectiveness of multiple interventions. (90)

However, utility instruments also have limitations. To appeal broadly, they include a common set of domains that most people value. This means that important domains can be underrepresented when used in specific conditions. (7,8) As a consequence, the sensitivity to capture specific treatment and condition effects may be reduced. (7,80)

Compared to the domains included in Neuro-QoL, at face value neither the SF6D nor HUI3 contain all domains considered relevant in describing HRQoL in persons with neurological conditions. For instance, fatigue has been identified as an important domain in neurological conditions; affecting 75% to 90% of MS patients. (8,91) The HUI3 does not directly evaluate fatigue and its omission is a major limitation of the instrument. The SF6D does evaluate vitality, which is moderately correlated with Neuro-QoL's fatigue domain (Rho = - 0.49) but is not a perfect substitute for fatigue.

However, despite the SF6D and HUI3 not including and thereby not *directly* evaluating a number of domains included in Neuro-QoL, both instruments were sensitive to most domains as evidenced by the moderate to strong correlations with Neuro-QoL domains (Tables 3.3 and 3.4). Thus, the SF6D and HUI3 appear to *indirectly* capture many of these important domains. For example, the HUI3's "within-the-skin" focus omits direct evaluation of social functioning; however, the HUI3 utility score is strongly correlated with Neuro-QoL's "ability to participate in social roles and activities" (Spearman's Rho 0.58) and moderately correlated with "satisfaction with social roles and activities" (Spearman's Rho 0.49). Similarly, the SF6D does not directly evaluate cognition but shows a similar strength of association with the Neuro-QoLs cognition domains as the HUI3, which does evaluate cognition. The SF6D is correlated with Neuro-QoL's cognition domains of general cognitive concerns (0.42) and executive function (0.44), compared to the HUI3's correlations of 0.46 and 0.49 respectively.

Due to a concern that limitations and omissions of important domains may limit the validity, interpretability and/or meaningfulness of utility scores produced, an emerging area of interest is the development of disease-specific utility measures to address the concern that generic preference-based measures are not sensitive to disease-specific improvements/detriments. (7,8,80) However, any potential gain in superior sensitivity in a disease-specific context is at the expense of comparability of utility values across conditions and instruments and, in some cases, insensitivity to other side effects and comorbidity. The introduction of condition-specific utility instruments is a welcome contribution to address some of the measurement shortcomings of generic utility instruments in specific condition contexts. However, more research is needed regarding

the full implications of expanding the utility instrument choices into condition-specific utility instruments. (80)

There are a number of potential limitations that may affect the results of this study. Two main, possible sources of error include: 1) the representativeness of the sample because it is from a non-random sample, and 2) the 20% non-completion rate suggesting a level of survey fatigue. The representativeness of the LINC sample has been compared to the Statistics Canada's Survey of Living with a Neurological Condition in Canada (SLNCC). (92) Generally, participants in the LINC study are slightly older, more educated, and have greater disutility than the individuals in the SLNCC. However, limited representativeness would impact all three instruments in a similar way and is unlikely to impact the conclusions of the study in any way. However, because the order of the instruments was not altered during the administration of the LINC survey, the high rate of non-completion suggests a potential loss of accuracy as individuals progressed through the survey. It is unclear how this loss of accuracy would impact study results and conclusions.

The cross-sectional nature of the data precluded the evaluation of the responsiveness of the HUI3 and SF6D to change. The responsiveness of an instrument relates to its ability to detect a change in HRQoL status over time when change *has* occurred. (49) Capturing change over time or change following the implementation of an intervention is valuable. Understanding how responsive the SF6D and HUI3 are in neurological conditions is an important next step in understanding the appropriateness of these instruments in neurological conditions and an area for future research.

This study adds to the growing body of evidence regarding the measurement of HRQoL in persons with neurological conditions. Generic utility instruments have a place in the evaluation of HRQoL when there is a need to summarize the impact of neurological conditions across multiple domains of HRQoL and if there is a desire to compare results to other populations or to an intervention. Given that the SF6D and HUI3 lack some domains relevant to neurological conditions, the use of a utility instrument in conjunction with a condition-specific instrument will enrich the description of HRQoL to ensure that condition relevant domains are captured.

3.5 CONCLUSIONS

The SF6D and HUI3 are both sensitive to variation in HRQoL domains relevant to persons with neurological conditions. The SF6D and HUI3 each measure a different array of HRQoL domains. Each, therefore, provides a complementary evaluation of the impact and burden of neurological conditions on everyday lives of Canadians. However, their evaluations of HRQoL in persons with neurological conditions are not fully comprehensive, because both instruments lack several domains important and relevant to those neurological conditions.

Table 3.1 Sociodemographic and clinical characteristics of the study population

Sample Characteristics (categorical variables)		n (%)
Sex	Male	244 (35)
	Female	444 (65)
Education	No high school	45 (7)
	High school	78 (12)
	Some post sec.	53 (8)
	Post secondary	491 (73)
Marital status	Married/common law	429 (62)
	Single/never married	151 (22)
	Widow/separated/divorced	107 (16)
Employment	Working	254 (36)
	Not working due to health	264 (37)
	Not working – other	193 (27)
Self-rated Health	Excellent	44(6)
	Very good	171(22)
	Good	304(39)
	Fair	194(25)
	Poor	57(7)
Health in past year	Health same or better	510 (66)
	Health declined	265 (34)

Sample Characteristics (continuous variables)	n	Mean (SD)	(95%CI)	Range
Age (mean)	690	50.82 (14.69)	(49.72, 51.92)	17 to 87
Co-morbid conditions (non-neuro) ^a	775	0.60 (0.74)	(0.55, 0.66)	0 to 3
No. neurological conditions/per person	766	1.53 (0.83)	(1.47, 1.59)	1 to 7
Time since Diagnosis (years)	669	17.34 (15.12)	(16.19, 18.49)	0.10 to 71.10
Time since Symptoms (years)	641	21.43 (15.42)	(20.24, 22.64)	0.10 to 71.10

Frequency of Neurological Conditions in the sample ^b			
	n		n
Migraine	159	Multiple sclerosis	183
Epilepsy	117	Parkinson's	147
Spina bifida	40	Muscular dystrophy	65
Hydrocephalus	36	ALS	26
Cerebral Palsy	14	Alzheimer's	11
Tourette's	8	Huntington's	7
Brain injury	108	Dystonia	43
Spinal cord injury	47	Brain/spinal cord tumor	24
Stroke	38		

a. Non-neurological comorbid conditions include: diabetes, heart disease and depression.

b. Neurological condition diagnosis categories are not mutually exclusive - an individual can have more than one diagnosis.

Table 3.2 Health-related quality of life of the study population

Utility Measures	n	Mean (SD)	(95%CI)	Range
SF6D utility score	766	0.62 (0.11)	(0.62, 0.63)	0.30 to 1.00
HUI3 utility score	766	0.47 (0.30)	(0.45, 0.50)	-0.28 to 1.00
Neuro-QoL Scales	n	Mean (SD)	(95%CI)	Range
Upper extremity ^a (GPR)	763	43.04 (10.74)	(42.27, 43.80)	12.8 to 53.8
Lower extremity ^a (GPR)	762	44.12 (11.44)	(43.29, 44.91)	16.5 to 58.6
Cognition – general concerns ^a (GPR)	753	41.82 (9.71)	(41.14, 42.52)	20.0 to 59.3
Cognition – executive function ^a (GPR)	747	43.45 (10.47)	(42.70, 44.20)	13.1 to 57.6
Fatigue ^b (NCS)	763	51.05 (8.62)	(50.44, 51.66)	29.5 to 74.1
Sleep disturbance ^b (NCS)	760	53.38 (9.14)	(52.74, 54.04)	32.0 to 84.2
Emotional/behavioral dyscontrol ^b (NCS)	748	49.35 (9.8)	(48.64, 50.06)	32.2 to 82.6
Depression ^b (NCS)	746	48.88 (8.66)	(48.25, 49.49)	36.9 to 75.0
Anxiety ^b (NCS)	749	50.35 (8.60)	(49.73, 51.00)	36.4 to 76.8
Positive affect and well-being ^a (GPR)	746	50.94 (8.61)	(50.41, 51.68)	26.3 to 68.0
Stigma ^a (NCS)	742	52.40 (7.87)	(51.73, 52.88)	39.2 to 81.5
Ability to participate in social roles/activities ^a (GPR)	719	45.88 (7.69)	(45.34, 52.47)	24.1 to 60.2
Satisfaction with social roles ^a (GPR)	719	44.28 (5.79)	(43.86, 44.72)	28.4 to 60.5

a. Higher Neuro-QoL score indicate *better* self-reported health.

b. Higher Neuro-QoL score indicate *worse* self-reported health

GPR = Standardized scale references a US general population

NCS = Standardized scale references a neurological clinical sample reference

Table 3.3 HUI3 attribute-to-attribute correlations with Neuro-QoL Scale dimensions (Spearman's Rho)

Neuro-QoL	HUI3 Utility Score	HUI3 (multi-attribute scores)							
		Ambulation	Dexterity	Cognition	Emotion	Pain	Speech	Vision	Hearing
Upper extremity ^a	0.50	0.54	0.73	0.03	0.02	0.22	0.26	0.16	0.02
Lower extremity ^a	0.61	0.83	0.54	-0.01	0.08	0.31	0.19	0.11	0.00
Cognition – general concerns ^a	0.46	-0.05	0.10	0.75	0.40	0.27	0.25	0.06	0.19
Cognition – executive function ^a	0.49	0.09	0.14	0.66	0.36	0.24	0.30	0.07	0.18
Fatigue ^b	-0.51	-0.21	-0.19	-0.33	-0.52	-0.43	-0.12	0.02	-0.09
Sleep disturbance ^b	-0.45	-0.10	-0.10	-0.36	-0.51	-0.45	-0.13	0.05	-0.05
Emotional & behavioral dyscontrol ^b	-0.39	0.02	0.02	-0.38	-0.49	-0.19	-0.10	0.06	-0.06
Depression ^b	-0.34	-0.05	-0.02	-0.34	-0.70	-0.27	-0.10	0.07	-0.06
Anxiety ^b	-0.30	0.03	-0.01	-0.39	-0.53	-0.24	-0.14	0.01	-0.10
Positive affect and well-being ^a	0.39	0.10	0.06	0.28	0.72	0.27	0.08	-0.07	0.05
Stigma ^a	-0.41	-0.20	-0.19	-0.22	-0.42	-0.29	-0.18	0.03	-0.04
Ability to participate in social roles ^a	0.58	0.32	0.27	0.35	0.50	0.37	0.20	0.01	0.07
Satisfaction with social roles ^a	0.49	0.29	0.23	0.25	0.49	0.32	0.17	-0.04	0.02

- a. Higher Neuro-QoL score indicate *better* self-reported health.
b. Higher Neuro-QoL score indicate *worse* self-reported health
Red correlations = strongly correlated (> 0.50)
Blue = moderately correlated (0.35 < Rho < 0.50)
Black = low correlation (< 0.35)

Table 3.4 SF6D attribute-to-attribute correlations with Neuro-QoL Scale dimensions (Spearman's Rho)

Neuro-QoL	SF6D Utility Score	SF6D (preference-weighted domain scores)					
		Physical Function	Role limitation	Social function	Vitality	Mental function	Pain
Upper extremity ^a	0.34	0.62	0.18	0.22	0.13	0.01	0.22
Lower extremity ^a	0.43	0.71	0.31	0.28	0.20	0.01	0.29
Cognition – general concerns ^a	0.42	0.08	0.29	0.42	0.29	0.39	0.28
Cognition – executive function ^a	0.44	0.17	0.29	0.44	0.25	0.36	0.29
Fatigue ^b	-0.66	-0.28	-0.38	-0.65	-0.49	-0.41	-0.46
Sleep disturbance ^b	-0.62	-0.19	-0.33	-0.55	-0.29	-0.49	-0.52
Emotional & behavioral dyscontrol ^b	-0.61	-0.02	-0.24	-0.39	-0.18	-0.49	-0.25
Depression ^b	-0.55	-0.15	-0.31	-0.55	-0.35	-0.64	-0.34
Anxiety ^b	-0.41	-0.09	-0.30	-0.49	-0.25	-0.66	-0.29
Positive affect and well-being ^a	0.57	0.16	0.30	0.52	0.37	0.55	0.30
Stigma ^a	-0.53	-0.27	-0.29	-0.53	-0.24	-0.37	-0.33
Ability to participate in social roles/activities ^a	0.67	0.35	0.39	0.69	0.39	0.38	0.43
Satisfaction with social roles ^a	0.60	0.31	0.33	0.61	0.36	0.37	0.36

a. Higher Neuro-QoL score indicate *better* self-reported health.

b. Higher Neuro-QoL score indicate *worse* self-reported health

Red correlations = strongly correlated (> 0.50)

Blue = moderately correlated (0.35 < Rho < 0.50)

Black = low correlation (< 0.35)

Table 3.5 Mean utility scores by normal, mild and moderate to severe impairment of Neuro-QoL domains

Neuro-QoL Scale Domain ^a		n (%)	Neuro-QoL ^b T-score(95%CI)	SF6D Mean (95%CI)	HUI3 Mean (95%CI)
Upper extremity (GPR)	Normal	322 (42)	53.8 (53.8, 53.8)	0.66 (0.65, 0.67)	0.61 (0.58, 0.64)
	Mild	134 (18)	42.5 (42.3, 42.8)	0.64 (0.62, 0.65)	0.52 (0.47, 0.57)
	Moderate & severe	307 (40)	32.0 (31.2, 32.7)	0.58 (0.57, 0.59)	0.31 (0.28, 0.34)
Lower extremity (GPR)	Normal	382 (50)	53.5 (52.9, 54.0)	0.67 (0.66, 0.68)	0.63 (0.61, 0.66)
	Mild	96 (13)	42.5 (42.3, 42.7)	0.60 (0.58, 0.61)	0.47 (0.42, 0.52)
	Moderate & severe	284 (37)	32.0 (31.2, 32.9)	0.57 (0.56, 0.58)	0.26 (0.23, 0.29)
Cognition – general concerns (GPR)	Normal	332 (44)	52.4 (51.8, 53.1)	0.68 (0.66, 0.69)	0.61 (0.58, 0.65)
	Mild	162 (22)	42.5 (42.3, 42.7)	0.63 (0.62, 0.65)	0.51 (0.47, 0.56)
	Moderate & severe	332 (44)	33.2 (32.7, 33.8)	0.58 (0.57, 0.59)	0.34 (0.31, 0.37)
Cognition-executive function (GPR)	Normal	314 (42)	53.7 (53.2, 54.2)	0.68 (0.67, 0.69)	0.62 (0.59, 0.65)
	Mild	150 (20)	42.4 (42.1, 42.7)	0.62 (0.60, 0.63)	0.51 (0.47, 0.54)
	Moderate & severe	283 (38)	32.7 (32.0, 33.3)	0.57 (0.56, 0.58)	0.30 (0.26, 0.33)
Fatigue (NCS)	Normal	507 (66)	46.3 (45.8, 46.8)	0.66 (0.66, 0.67)	0.56 (0.54, 0.59)
	Mild	166 (22)	57.8 (57.6, 58.1)	0.57 (0.55, 0.58)	0.37 (0.33, 0.41)
	Moderate & severe	90 (12)	65.4 (64.5, 66.2)	0.50 (0.49, 0.52)	0.17 (0.12, 0.22)
Sleep disturbance (NCS)	Normal	431 (57)	47.1 (46.5, 47.6)	0.67 (0.66, 0.68)	0.57 (0.54, 0.60)
	Mild	146 (19)	57.4 (57.2, 57.6)	0.59 (0.57, 0.60)	0.41 (0.37, 0.46)
	Moderate & severe	183 (24)	65.1 (64.4, 65.7)	0.54 (0.53, 0.55)	0.29 (0.25, 0.34)
Emotional/behavioral dyscontrol (NCS)	Normal	541 (72)	44.7 (44.2, 45.3)	0.65 (0.64, 0.66)	0.52 (0.49, 0.54)
	Mild	112 (15)	57.7 (57.5, 58.0)	0.59 (0.57, 0.60)	0.40 (0.35, 0.46)
	Moderate & severe	95 (13)	65.8 (64.9, 66.8)	0.54 (0.52, 0.56)	0.30 (0.25, 0.35)
Depression (NCS)	Normal	543 (73)	44.9 (44.3, 45.4)	0.66 (0.65, 0.67)	0.53 (0.50, 0.55)
	Mild	135 (18)	57.3 (57.0, 57.5)	0.56 (0.54, 0.57)	0.38 (0.33, 0.42)
	Moderate & severe	68 (9)	64.3 (63.4, 65.3)	0.51 (0.50, 0.53)	0.23 (0.17, 0.30)
Anxiety (NCS)	Normal	517 (69)	46.0 (45.5, 46.5)	0.67 (0.65, 0.67)	0.53 (0.51, 0.56)
	Mid	137 (18)	57.2 (57.0, 57.4)	0.57 (0.56, 0.59)	0.38 (0.33, 0.42)
	Moderate & severe	95 (13)	64.2 (63.4, 65.0)	0.52 (0.50, 0.53)	0.27 (0.22, 0.33)
Positive affect & wellbeing (GPR)	Normal	553 (74)	54.6 (54.1, 55.1)	0.65 (0.64, 0.66)	0.53 (0.50, 0.55)
	Mild	128 (17)	42.3 (42.1, 42.6)	0.56 (0.54, 0.57)	0.34 (0.29, 0.39)
	Moderate & severe	65 (9)	36.6 (35.8, 37.5)	0.52 (0.50, 0.54)	0.28 (0.21, 0.35)
Stigma (NCS)	Normal	441 (59)	47.4 (46.9, 47.9)	0.66 (0.65, 0.67)	0.56 (0.53, 0.59)
	Mild	177 (24)	57.2 (57.0, 57.4)	0.58 (0.57, 0.59)	0.40 (0.35, 0.44)
	Moderate & severe	124 (17)	63.4 (62.7, 64.1)	0.55 (0.53, 0.56)	0.28 (0.23, 0.32)
Ability to participate in social roles (GPR)	Normal	349 (49)	52.1 (51.4, 52.7)	0.69 (0.68, 0.70)	0.55 (0.53, 0.57)
	Mild	211 (29)	42.6 (42.4, 42.8)	0.59 (0.58, 0.60)	0.43 (0.40, 0.47)
	Moderate & severe	159 (22)	36.7 (36.3, 37.1)	0.53 (0.52, 0.54)	0.20 (0.19, 0.24)
Satisfaction with social roles and activities (GPR)	Normal	234 (33)	50.6 (49.9, 51.2)	0.70 (0.68, 0.71)	0.64 (0.60, 0.67)
	Mild	362 (50)	42.6 (42.5, 42.7)	0.61 (0.60, 0.62)	0.45 (0.42, 0.47)
	Moderate & severe	123 (17)	37.2 (36.7, 37.7)	0.53 (0.51, 0.54)	0.23 (0.20, 0.30)

a. Difference between severity level statistically significant for all comparisons at p-value <0.001

b. Neuro-QoL T-scores (0-100): mean = 50, standard deviation (SD) = 10.

Shaded domains: High Neuro-QoL scores indicate worse (undesirable) self-reported health

Unshaded domains: High Neuro-QoL scores indicate better (desirable) self-reported health

GPR Neuro-QoL scores standardized to a general population reference (GPR)

NCS Neuro-QoL scores standardized to a neurological clinical sample reference (NCS)

CHAPTER 4

(Manuscript)

Discordance in Utility Measurement in Canadians
with Neurological Conditions:

A Comparison of the SF6D and HUI3

ABSTRACT

BACKGROUND

Health related quality of life (HRQoL) is an important outcome in persons with neurological conditions as it contributes to a greater understanding of disease consequences and treatment effects. Differences in the way preference-based HRQoL (utility) instruments define, describe and value HRQoL affects the validity, interpretability, comparability, meaningfulness, and agreement of utilities generated by these instruments. Although the performance of these instruments has been studied before, they have not been evaluated in a diverse sample of neurological conditions.

OBJECTIVE

The study examined the extent of agreement / discordance between the SF6D and HUI3 in Canadians with neurological conditions and how it varied by participant and impairment characteristics.

METHODS

The study analyzed cross-sectional population survey data collected as part of the LINC Study (Living with the Impact of a Neurological Condition). Self-reported data was collected on the burden and impact of neurological conditions on participant's everyday lives. Agreement was explored comparing utility distributions, paired t-tests of the means, Shearman's Rho correlations, intra-class correlations (ICC), and Bland Altman plots. Associations between participant characteristics and utility differences were assessed using multiple regression (ordinary least squares (OLS)) models.

RESULTS

The data of 776 individuals with scores for both the SF6D and HUI3 was analyzed. The results of this analysis showed poor agreement between the two instruments. Discordance between the SF6D and HUI3 was substantial, with a mean utility difference of 0.15 (95% CI 0.13, 0.17), and an ICC coefficient of 0.41. The Bland Altman plot and regression analysis showed systematic variation in utility difference associated with HRQoL status. The pattern of discordance did not vary by participant and impairment characteristics.

CONCLUSIONS

Despite sharing a common purpose, the SF6D and HUI3 differ in many respects resulting in discordant evaluations of HRQoL in persons with neurological conditions. The substantial, clinically important differences in utility estimates between the SF6D and HUI3 therefore cast doubt on whether the utility estimates are directly comparable or universally valid.

4.1 INTRODUCTION

Neurological conditions are a significant cause of disability and disease burden in Canada. Although they are a diverse group of conditions, varied in their presentation, progressiveness, time course and sequelae; collectively, they are marked by high individual, caregiver, societal, and economic burden. (1,9) The impact of neurological conditions extends beyond physical and cognitive symptoms and impairments, and affects social relationships and roles with family, friends, and the wider community. (6) Very few neurological conditions are curable. As a result, the treatment emphasis is often on managing the impact of symptoms and the maintenance of health related quality of life. (3-6,8,10,67)

Health-related quality of life (HRQoL) is an important outcome to capture the “impact” of disabling conditions that are life-long and multidimensional. (7,8,68) Assessment of HRQoL contributes to a greater understanding of disease consequences and treatment effects on everyday life. (7,8) Among the options to quantify the “impact” of a neurological condition is the use of preference-based HRQoL instruments.

Preference-based HRQoL instruments (also commonly known as multi-attribute / utility / index instruments) are a unique class of generic instruments that evaluate HRQoL in terms of utility. Utility instruments describe the HRQoL of an individual on a theoretical zero to one scale; describing the relative value of health in comparison to perfect health (1.00) and death (0.00). Utility instruments are distinct from psychometric measures of HRQoL in that they incorporate preferences that reflect the desirability of a health state as “better” or “worse”. (12-15)

Preference-based instruments are comprised of two constituent parts: 1) a descriptive health classification system that defines and describes the health states of interest, and 2) a scoring system that assigns a value (preference) to the health states and converts the descriptive system into a summary utility score using a scoring algorithm. Evaluations of HRQoL are influenced by interactions between an instrument’s descriptive health classification system and scoring systems. (22,27)

The appeal of utility instruments is their ability to provide a single summary score across multiple health domains; a trait that lends itself well to assessments of HRQoL and

the need to balance effects across physical, mental, and socio-emotional domains. In this way utility instruments are able to simultaneously weigh the net gain (or loss) in HRQoL by summarizing improvements in some domains and detriments in others. This is important in conditions where the impact on HRQoL spans multiple domains or in assessing interventions that may provide benefit in some HRQoL domains at the expense of others. Additionally, utility instruments are promoted as instruments that, in theory, permit broad comparisons across populations and interventions because of their index (ratio) scale. (23) These features have facilitated the ready inclusion of utility instruments in population surveys and clinical applications across a broad spectrum of disease categories over the past two decades. (27)

However, utility instruments also have limitations. Even though these instruments are designed to measure the same concept and have a similar purpose, they vary in many respects in terms of their descriptive health classification systems and scoring systems. Comparative studies frequently yield differences in utility distributions, sensitivity, and responsiveness between instruments, irrespective of the instrument or population. (20,44)

The lack of agreement, or discordance, between utility instruments is widely documented in the literature. The degree of discordance is of significant concern given that utility instruments are recommended for use in economic evaluations (e.g. by the Canadian Agency for Drugs and Technologies in Health (CADTH) and the National Institute for Health and Clinical Excellence (NICE)). (8) The accumulating empirical evidence of discordance between these instruments challenges the universal validity of HRQoL evaluations required for decision-making and is a critical and unresolved issue. (27) The concern is that utility scores, and hence potential evaluations and decisions, may vary simply according to the choice of instrument. (16,30,31,44,66)

Discordance may be particularly pronounced in measuring utility in persons with neurological conditions due to the level of disability and the complexity and diversity of impairments within and between conditions. (3,7,30,74-76,78,93) Both the progression of a neurological condition, and the interventions designed to relieve symptoms and impairments, may affect HRQoL domains differentially and thereby complicate HRQoL assessment.

The purpose of this study was to explore agreement and discordance between the Short-Form 6D (SF6D) (26) and the Health Utilities Index – Mark 3 (HUI3) (22,23) in Canadians with neurological conditions. Comparisons of agreement between utility instruments in neurological conditions are limited. (20,36,43,94-96) No prior study has evaluated agreement / discordance between the SF6D and HUI3 in relation to the common symptoms and functional limitations shared by neurological conditions as a whole. Specifically, the objectives of this study were to:

1. Describe agreement and discordance between the SF6D and HUI3.
2. Describe how discordance varied by participant characteristics, type of neurological condition, and HRQoL domains relevant to persons with neurological conditions.

Meeting these objectives will help guide answers to questions such as: Do these instruments measure HRQoL in individuals with neurological conditions in a similar way? Are there characteristics common to individuals or to neurological conditions that are associated with measurement discordance? Is there evidence that an instrument may favour or disadvantage individuals who have particular characteristics?

4.2 METHODS

This study explored a number of descriptive, methodological and empirical differences between the SF6D and HUI3 in describing HRQoL in a sample of Canadians with neurological conditions. The data was drawn from the LINC Study (Living with the Impact of a Neurological Condition) and includes data that describes the health burden and impact on participation in everyday life of people with neurological conditions. (81)

4.2.1 Sample and data collection

A sample of 948 Canadians, 17 years of age or older, with at least one identified neurological condition participated in the cross-sectional population survey administered as part of the LINC Study. The sample was a non-random (convenience) sample and was recruited by the Neurological Health Charities Canada (NHCC) and NHCC member organizations. Recruitment strategies for the study included a variety of word-of mouth

invitations distributed through relevant organizations, known databases, and registries of people with neurological conditions

Participants completed a comprehensive survey on the burden and impact of neurological conditions on their everyday lives between September 14, 2011 and July 1, 2012. The majority of respondents (89%) completed the survey online using Opinio (83), a web-based survey application. The remaining respondents completed the survey on paper (8%) or by telephone (3%). Self-reported data on diagnosis and socio-demographics such as: age, sex, marital status, employment status, and education were collected. Clinical information included time since diagnosis, self-rated health, indication of a decline in health in past year, and presence of other co-morbid conditions (diabetes, heart disease, and depression). Data was exported from Opinio, examined and cleaned using Excel (84), and derived variables were calculated using SAS (85) for standard measures, scales and indices. Missing data was handled as per each instrument's standard protocol.

This study analyzed data of the 776 individuals who had utility scores for both the SF6D and HUI3. From the original sample of 948, 172 individuals were excluded using pairwise deletions due to 95 missing SF6D and/or 134 missing HUI3 scores. There were some differences between survey completers and non-completers; those with missing HUI3 and/or SF6D utility scores were more likely to have an early-onset/traumatic condition (degenerative conditions were associated with missing HUI3 values only). In addition, individuals missing HUI3 scores were more likely to have greater anxiety, lower self-efficacy, and were more likely to need help with self-care.

4.2.2 Measures of HRQoL

The LINC population survey was constructed using a series of validated scales and standardized instruments selected to collect data about the impact of neurological conditions. The analysis used data from the Short-Form 6D (SF6D), ⁽²⁶⁾ Health Utilities Index – Mark 3 (HUI3), (22,23) and Quality of Life in Neurological Disorders (Neuro-QoL) measures. (10,11)

The SF6D and HUI3 are generic, indirect, preference-based multi-attribute health state classification and utility scoring systems. Each instrument can provide a profile of health describing the extent of disability by HRQoL domain and also summarize HRQoL

as a single utility-based index score. (22,26) However, while the two instruments have been designed to measure the same concept, “HRQoL”, they vary widely in terms of their operationalization of health, the number and type of HRQoL domains and domain levels, preference elicitation methods, and scoring function. For example, one of the most notable differences between the SF6D and HUI3 is in the operationalization of HRQoL. The HUI3 health classification system is based on a narrower “within-the-skin” definition of health. (32,41,42) It focuses on actual impairments and defines health status in terms of capacity (29) (i.e. vision, speech, and ambulation), but omits social interaction as considered “outside-the-skin”. Whereas the SF6D, based on the SF-36v2, has a broader definition of health and, in addition to measuring physical and mental components of health, also measures emotional role and social functioning. (27)

Neuro-QoL is a set of HRQoL measures applicable for people with common neurological conditions. (67) Neuro-QoL was developed as a clinically relevant and psychometrically robust QOL assessment tool for neurologic disorders, to enable clinical researchers to compare the QOL impact of different interventions within and across various neurological conditions. (10,11)

Health Utilities Index – Mark 3(HUI3) (www.healthutilities.com) (22,23)

The HUI3 describes 8 domains of HRQoL: vision, hearing, speech, dexterity, ambulation, cognition, emotion, and pain. Each domain has 5-6 levels of function. The instrument describes 972,000 unique health states. Health state valuation used standard gamble from a sample of 504 Canadians. HUI3 utility scores can range from -0.36 to 1.00. The HUI3 utility score is calculated using the multi-attribute scoring function and Canadian sample preference weights provided by Health Utilities Incorporated. The HUI3 has widespread use in clinical and general applications, has been included in major population health surveys in Canada since 1990 and is one of the most frequently used tools in economic analysis. (18,22) (*Further details in Appendix A*)

Short-Form-6D (SF6D) (derived from SF-36v2) (www.SF36.org) (26)

The SF6D describes 6 domains of HRQoL: physical functioning, role limitations (physical and emotional), bodily pain, vitality, social functioning, and mental health. Each domain has 1-3 items, which have 4 to 6 levels of function. The instrument describes 18,000 unique health states. Health state valuation used standard gamble from a sample of 836 individuals in the UK. SF-6D utility scores can range from 0.30 to 1.00. The SF6D utility score and health domains are calculated from 11 of the 36 items of the SF-36v2 completed in the survey, based on the SF6D preference-based algorithm and UK preference weights provided to the author by Brazier and colleagues. (26,86-88) The SF6D is more widely used in Europe and unlike other instruments that are conceived and developed specifically as preference-based tools, the SF6D is unique in that its utilities are derived using an algorithm developed by Braizer and colleagues from 11 of the 36 items of the well-established psychometric instrument – the Short Form-36 (SF-36v2). (19,26,27,45) Given that the SF-36v2 is the most widely used psychometric measure of HRQoL, the development of the algorithm to convert existing and future SF-36 results into utility scores has been seen as a valuable addition to utility measurement. (12,35) *(Further details in Appendix B)*

SF6D and HUI3 HRQoL domains and number of domain levels

SF6D Domains (No. of domain levels*)	HUI3 Domains (No. of domain levels*)
Physical functioning (6 levels)	Ambulation (6 levels)
Role limitation (4 levels)	Dexterity (6 levels)
Social functioning (5 levels)	Cognition (6 levels)
Mental health (5 levels)	Emotion (5 levels)
Pain (6 levels)	Pain (5 levels)
Vitality (5 levels)	Vision (6 levels)
	Speech (5 levels)
	Hearing (6 levels)

* In both instruments, domain levels are based on categories of functioning (Likert type scale). Domain level 1 represents full functioning in the domain. Functioning declines with increasing domain levels, where levels 4 to 6 represent severe HRQoL impairment.

Quality of Life in Neurological Disorders measures (Neuro-QoL) (www.neuroqol.org)
(10,11)

Neuro-QoL is a condition-specific outcome measure designed to generate clinically relevant and psychometrically robust, health-related quality of life assessment data, across a broad spectrum of symptoms and issues relevant to neurological conditions. It was developed in response to a lack of specificity in generic instruments; a lack of a common condition-specific instrument used in neurological conditions; and a desire to better capture a person's experience of neurological disease symptoms, treatment side effects, functioning and well-being. (10,11) While Neuro-QoL is labeled as a health related quality of life measure, it does not provide a summary measure of HRQoL. Rather, Neuro-QoL contains a series of separate, condition-specific psychometric scales of HRQoL domains relevant to many neurological disorders. Thirteen of the Neuro-QoL short form scales were included in the LINC study: upper extremity function, lower extremity function, cognition (general and executive function), depression, anxiety, fatigue, sleep disturbance, emotional and behavioral dyscontrol, positive affect and wellbeing, ability to participate in social roles and activities, stigma, and satisfaction with social roles and activities. Each scale has 8-9 items describing the domain. Raw scores for each scale were converted to T-scores based on the scoring tables in the Neuro-QoL user manual. T-scores have a mean of 50 and standard deviation of 10. The scores are either standardized in reference to the US general population or a clinical population. Higher scores in a Neuro-QoL scale indicate that more of a domain is measured. For example, a high fatigue score indicates worse HRQoL as it relates to fatigue than a lower score, whereas, a high "lower extremity" score indicates better HRQoL in this domain. A 0.5 SD change in T-score is considered a clinically relevant difference. (89) *(Further details in Appendix C)*

4.2.3 Statistical Analysis

The analysis for this study involved a number of steps. First, descriptive statistics were used to characterize the sample in terms of utility scores and participant characteristics. Means and 95% confidence intervals (95% CI) were reported for continuous variables and frequencies and proportions (%) for categorical variables. In

order to meet the study objective, to describe discordance between the SF6D and HUI3, and how it varied by participant characteristics, the analyses: i) quantified discordance between the SF6D and HUI3, ii) assessed the relationship between SF6D and HUI3 attributes, iii) identified the presence of ceiling and floor effects, and iv) explored characteristics affecting utility difference between the SF6D and HUI3. Differences in the SF6D and HUI3 descriptive health classifications systems were also explored to evaluate their contribution to discordance.

Variables of interest

Throughout the analysis, the response variable of interest was discordance calculated as utility difference (dUTY): $dUTY_i = SF6D_i - HUI3_i$. Explanatory covariates included a range of participant characteristics organized into three categories: 1) person characteristics, 2) type of neurological condition, and 3) HRQoL domains relevant to persons with neurological conditions based on the Neuro-QoL measures. Neuro-QoL is a set of HRQoL measures designed to evaluate multiple HRQoL concerns common to, and applicable for, many neurological conditions. For the purpose of this study, Neuro-QoL was used to represent HRQoL domains relevant and important to the evaluation of HRQoL in persons with neurological conditions. (67)

Participant Characteristics (explanatory covariates)

<i>Person characteristics</i>	<i>Type of Neurological Condition*</i>	<i>Neuro-QoL HRQoL domains</i>
Age Sex Education Employment status Marital status, Co-morbid conditions (non-neuro) Time since diagnosis Change in health status in past year Self-rated health	Episodic Degenerative Traumatic / sudden-onset Congenital /early-onset * These categories are not exhaustive or mutually exclusive as some individuals or diagnoses may be included in more than one category.	Upper Extremity Lower extremity Cognition-general concerns Cognition- executive function Fatigue Sleep disturbance Depression Anxiety Emotional behavioural dyscontrol Positive affect and wellbeing Stigma Ability to participate in social roles Satisfaction with social roles

Important Differences

Throughout the analysis, difference in utility was assessed for both its statistical significance ($p\text{-value} \leq 0.05$) and its magnitude, or effect size. Minimally important differences (MID) vary from instrument to instrument and study to study. For the purpose of this study the frequently cited minimally important utility difference of 0.03 was adopted as a threshold of agreement / discordance between the SF6D and HUI3. (22,95,97,98)

Correlational analyses were evaluated based on Guyatt and colleagues (1993) suggested correlation classifications: where greater than 0.50 represents a strong relationship; correlations between 0.35 and 0.50 represent a moderate relationship; and below 0.35, a marginal relationship between domains. (14)

4.2.4 Analysis Steps

The analysis involved 4 steps to describe discordance between the SF6D and HUI3 overall and how it varied by participant characteristics, type of neurological condition, and HRQoL domains relevant to persons with neurological conditions.

Step 1. Quantify discordance between the SF6D and HUI3

The extent of agreement/discordance between the SF6D and HUI3 was explored using the commonly cited measures of agreement. Descriptive statistics and distributions of summery utility scores were examined. Within-subject differences in utility score means were tested using paired t-tests for person characteristics and type of neurological condition covariates. Agreement was assessed using an intra-class correlation (ICC) based on a two-way mixed model, where a coefficient below 0.40 was considered poor agreement and between 0.41 and 0.75 moderate agreement. (99,100) Because the ICC is considered a poor assessment of agreement, due to its tendency to be influenced by the heterogeneity of the sample, between-instrument discordance and its relationship with overall HRQoL (represented by mean utility) was explored using a Bland-Altman plot. (54)

The Bland-Altman plot graphically depicts the mean measurement bias and shows if systematic variation in discordance is present. The plot was constructed by plotting the data points:

$$S(y, x) = (SF6D - HUI3), \frac{SF6D + HUI3}{2}$$

The y-axis represents the difference between utilities and the x-axis, the mean of the two utilities. If the instruments generally agree across the range of HRQoL, the data points will hover around zero, scattered arbitrarily with no consistent pattern. If discordance is evident the mean difference, or “measurement bias”, will deviate away from zero. Expected variation for paired observations was estimated as the 95% limits of agreement (mean \pm 1.96 SD). Widely spaced limits of agreement suggest serious levels of disagreement. (59)

Step 2. Assess the relationship between SF6D and HUI3 HRQoL domains

The similarities/differences between the SF6D and HUI3 descriptive health classification systems (e.g. measuring similar HRQoL constructs and including similar HRQoL domains) were compared as a potential source of discordance. This was achieved by assessing the strength of the relationship between domains of the SF6D to domains of the HUI3 using a Spearman correlation matrix. Strong correlations were expected between “similar” domains such as physical functioning (ambulation/dexterity), mental (emotional) health, and pain.

Step 3. Identify presence of ceiling and floor effects

Ceiling and floor effects are a methodological issue that restricts the variance of an instrument and can affect agreement between instruments; particularly at ends of the utility distributions. A common way to measure the presence of a ceiling or floor effect in cross-sectional data is to quantify the proportion of respondents through the range of utility scores. Ceiling and floor effects are identified if an instrument shows a clustering of scores in the upper or lower ends of utility distribution. (49)

To identify ceiling effects, the proportions of utility scores at the upper limit of the utility distribution in both instruments between 0.90 and 1.00 were reviewed. Floor effect detection was complicated by the SF6D lower utility limit of 0.30. Floor effects in the SF6D were considered in utilities between 0.30 and 0.40; and between -0.28 and -0.18 in the HUI3. For the purpose of this study, a ceiling or floor effect was considered to be present if more than 10% of scores fell into the upper or lower limit of an instrument's distribution. To assess the contribution of individual domains to the ceiling and/or floor effects observed, the proportion of responses across HRQoL domain levels were also examined within the ceiling and floor ranges.

Step 4. Explore Factors affecting utility difference between the SF6D and HUI3

The association of participant characteristics, type of neurological condition, and Neuro-QoL domains with discordance was explored using multiple regression (ordinary least square (OLS)). Utility difference (dUTY) was the dependent variable and is calculated as:

$$dUTY_i = SF6D_i - HUI3_i$$

In preliminary analysis of the data, a Bland-Altman plot showed that utility difference was strongly associated with overall HRQoL (mean utility). This association was mainly linear but the slope of the association changed direction abruptly between mean utility values of 0.7 and 0.8. For the current analyses, mean utility was added to the model to control for these associations and spline modeling techniques were also employed.

Spline modeling is a useful technique when linearity assumptions are violated. As a more flexible modeling technique, it permitted the use of all information available in the mean utility variable. (See Appendix E) Spline modeling estimated the association between mean utility and utility difference (dUTY) in a piecewise linear fashion. Regression analysis assessed whether both the level and pattern of the piecewise association varied by the explanatory

covariates (person, type of neurological condition, and HRQoL domain characteristics). (101)

To mirror the pattern of discordance associated with mean utility observed in the Bland-Altman plot, mean utility was divided in to three linear splines. The splines were established with two knots located at mean utilities: 0.30 and 0.77. “Knots” are the points where the splines intersect and give the linear segments freedom to bend and more closely follow the data. Because the utility ranges of SF6D and HUI3 are different, the knot at 0.30 was chosen to reflect the SF6D’s end range of 0.30. The knot at 0.77 was chosen to capture the change in direction in the slope of discordance observed in the Bland-Altman plot. (See Appendix F)

A base model, including only the splines of mean utility, was first estimated. Once estimated, and with the effect of mean utility controlled for, individual associations of the explanatory covariates on discordance were explored. Each multiple regression spline model was in the form of:

$$\widehat{dUTY} = \alpha + f(\beta_{mean\ utility}) + \beta_j x_j$$

where f is the spline function to estimate the data using the 3 piecewise linear splines of mean utility, and β_j represents the estimated discordance associated with a particular covariate or unit change in a covariate, and x_j the covariate of interest. The main effect of covariates, β_j , measured changes in the intercept and reflect vertical translations of the slope of discordance associated with the selected characteristic. Each model was re-run controlling for age and sex. A final model contained all 13 Neuro-QoL variables to assess the effect on discordance, while taking into account all potential Neuro-QoL domains. In addition, interactions between the explanatory covariates and the liner splines were explored and assessed for statistical significance. Coefficients for these interactions assessed whether the slope of discordance varied by the explanatory covariates.

All models were assessed for fit statistics such as the F statistic, p-value and the percentage of variance (R^2). Regression estimates were evaluated for

statistical significance ($p\text{-value} \leq 0.05$) and magnitude, or effect size.

Conventional regression diagnostics were performed on the baseline discordance model. All study data was analyzed using STATA 12 (StataCorp). (82)

4.3 RESULTS

Characteristics of the sample

The data of 776 individuals who had scores for both the SF6D and HUI3 were analyzed. Characteristics of the sample are presented in Table 4.1. The sample had a mean age of 50.8 years (95% CI 49.7, 51.9), were predominantly female (64.5%), were highly educated, and a high proportion of the sample (64%) did not work due to health or other reasons. Seventeen neurological diagnoses were represented in the sample (Table 4.2). Participants reported on average 1.5 (SD 0.83) neurological conditions per person. The impact of neurological conditions on HRQoL was high, evidenced by low utility scores. (Table 4.2).

Descriptive statistics of the SF6D and HUI3

The HUI3 had a mean utility of 0.47 (95% CI 0.45, 0.49) with a range of -0.28 to 1.00; whereas the SF6D had a mean utility of 0.62 (95% CI 0.62, 0.63) and a narrower range of 0.30 to 1.00. The means and medians for each instrument were identical. There were marked differences between the utility distributions as shown in Figure 4.1 and in utility scores across participant characteristics shown in Tables 4.1 and 4.2. Individuals were dispersed fairly evenly through the range of utility for the HUI3 but not in the SF6D. The Shapiro-Wilk test for normality confirms neither instrument's utility scores were normally distributed ($p < 0.001$).

Discordance between the SF6D and HUI3

The mean difference in utility was 0.15 (95% CI 0.13, 0.17). Substantial utility discrepancies across all participant characteristics were confirmed by paired t-tests (Tables 4.1 and 4.2). Despite a strong linear association between the SF6D and HUI3 utility scores (Pearson's $r=0.62$, $p\text{-value} < 0.0001$ and Spearman's $\rho = 0.61$, $p\text{-value} < 0.0001$) (Figure 4.2), the intraclass correlation (ICC) demonstrated *marginal* agreement, with a coefficient

of 0.41 (95% CI 0.35 to 0.46) suggesting only 41% of the variation in utility was due to between-subject differences.

Bland-Altman plot (Figure 4.3) analysis also highlighted a lack of agreement between the two instruments. The 95% limits of agreement were wide (-0.34 to 0.65) with potential variation in utility (0.99) equivalent to the theoretical 0.00-1.00 utility score range. On average, the SF6D provided a utility score 0.15 higher than the HUI3. However, discordance varied systematically with HRQoL status (mean utility). The magnitude of disagreement was more extreme at lower levels of HRQoL where the SF6D provided a higher utility than the HUI3; a trend that crossed over at a utility of approximately 0.65 following which the HUI3 provided a higher utility score than the SF6D.

Relationship between SF6D and HUI3 HRQoL domains

Table 4.3 shows the strength of relationship between SF6D and HUI3 HRQoL domains. Only 4 of the possible 48 intercorrelations showed strong positive correlations ($\rho > 0.50$) and were between domains considered most similar: physical function (ambulation and dexterity); mental (emotional) health, and pain. Beyond these associations, the HUI3's emotion and pain attributes were moderately associated with the SF6D's role limitations, social functioning and vitality attributes. There were only marginal correlations, ranging from 0.02 to 0.20, between the HUI3 "sense" attributes (vision, hearing, and speech) and the SF6D's attributes. Pain was most highly correlated at 0.72.

Ceiling and floor effects

Ceiling and floor effects were identified by reviewing the distribution of participants through the range of utility values for each instrument. The HUI3 showed a mild ceiling effect with 10.2% of participants between a utility of 0.9 to 1.00 in the HUI3 compared to only 1.8% in the SF6D. Neither instrument showed a floor effect. Only 1.3% of participants fell between SF6D utility scores of 0.30 and 0.40; only 1% of participants had an HUI3 utility score between -0.28 and -0.18. However, there was an extreme clustering of SF6D scores (68.3%) between utilities of 0.50 to 0.69 and a clustering of HUI3 utility scores (45%) between utilities of 0.30 to 0.49.

A detailed inspection of the proportion of participants within SF6D and HUI3 HRQoL domains was undertaken where a clustering of utility scores was observed (Table 4.4). With respect to the ceiling effect seen in the HUI3, a high proportion of the study sample (ranging from 20.8% to 87%) was clustered within the HUI3's HRQoL domain level 1, where level 1 indicates no impairment. In contrast, a much smaller proportion of the sample (2.2% to 21.3%) was within level 1 of the SF6D domains.

With respect to the extreme clustering of SF6D utility scores between 0.50 and 0.69, inspection of the proportion of participants distributed through the SF6D domain levels (Table 4.4b) did not provide clarity as to the cause of this clustering. With the exception of the role participation and vitality domains, most participants in the sample were dispersed widely through the SF6D domain levels. Role participation showed a domain floor effect with 60.7% of sample reporting lowest functioning (level 4), and vitality was more heavily distributed in the lower levels of the vitality domain.

Characteristics associated with discordance between the SF6D and HUI3

Tables 4.5, 4.6 and 4.7 show the results of multiple regression analysis estimating the associations between the explanatory covariates and discordance (dUTY). The base model accounted for the systematic variation seen between discordance and mean utility in the Bland-Altman plot and represented the plot in regression form. Participant characteristics (Table 4.5), type of neurological condition (Table 4.6), and Neuro-QoL HRQoL domain (Table 4.7) covariates were placed in the base model one at a time to assess individual effects on discordance (dUTY).

The baseline discordance model predicted a mean discordance (dUTY) of 0.15 (95%CI 0.14, 0.16.) and HRQoL status was associated with 73% of the variation in discordance with an R^2 of 0.7324, F -test = 704.2, p -value 0.0000. The addition of explanatory covariates to the models did not alter the pattern of discordance predicted by the Bland-Altman plot and baseline regression model. The modest effect sizes of the regression β coefficients shown in Tables 4.5, 4.6 and 4.7 indicate that *change* in discordance by covariates varied only marginally beyond the discordance associated with HRQoL status (base model). Adding age and sex to the models did not alter the conclusions in anyway. The effect of participant characteristics on discordance was

marginal (Table 4.5), and type of neurological condition (Table 4.6) did not show statistically or clinically important differences (MID) (> 0.03) in discordance. The majority of Neuro-QoL covariates showed modest statistically and clinically important differences in discordance (Table 4.7).

Each model was re-run with interaction terms between explanatory covariates and the splines of mean utility. None were found to be significant, indicating that the explanatory covariates were only associated with modest variations in the magnitude of discordance; not a change in the *pattern* of discordance.

4.4 DISCUSSION

This study is the first head-to-head comparison of the SF6D and HUI3 in persons with diverse neurological conditions. The study evaluated whether the SF6D and HUI3 measured HRQoL in the same way by evaluating the extent of agreement / discordance between the SF6D and HUI3 and describing how it varied by participant and impairment characteristics.

The SF6D and HUI3 did not measure HRQoL in persons with neurological conditions in the same way. Substantial differences were seen in utility scores, utility distributions, and utility range. On average, the SF6D provided a utility score that was 0.15 higher than the HUI3. The SF6D and HUI3 did not agree in their evaluation of HRQoL and showed a lack of mapping between HRQoL domains. A key finding was that discordance between the SF6D and HUI3 varied systematically with overall HRQoL status.

The results of this study are consistent with the evidence from other comparative reviews of utility instruments that find: 1) different instruments provide different utility estimates in the same population, 2) the descriptive health classifications systems of the instruments are not equivalent, and 3) the choice of instrument could influence decision-making. (12,27,29,33,43,53) Utility scoring system differences are also known to impact discordance but evaluating their contribution to discordance was beyond the scope of this study.

Most respondents (92%) had utility differences greater than 0.03. The mean difference in utility (0.15) in this study was greater than prior comparison studies within

other diseases and in general populations studies. Utility difference ranged from minimal in general population studies to utility differences of 0.10 in studies involving specific conditions such as Rheumatoid arthritis, patients of hip arthroplasty, and kidney disease. (28,41,52,55,61) The exception is a study by Pickard and colleagues (2005) which compared the HUI3 and SF6D in stroke recovery and found a utility difference of 0.36 between the SF6D and HUI3 at baseline and a difference of 0.24 at 6 months. The utility scores for these patients were very low (e.g. at baseline, the SF6D mean (SD) utility was 0.55 (0.09) and the HUI3 mean (SD) utility was 0.19 (0.30)). (30) The substantial discordance associated with very low utility scores found by Pickard and colleagues is consistent with the current study's findings that suggest greater discordance occurs in samples with lower HRQoL status.

Additionally, HUI3 utility in persons affected by MS, stroke, and Parkinson's included in the LINC sample was much lower than population survey reports based on the Canadian Community Health Survey (CCHS).

Condition	LINC n	LINC SF6D Mean (95%CI)	LINC HUI3 Mean (95%CI)	LINC Mean Difference	Canadian Sample Utility (HUI3 only)
Stroke	38	0.59 (0.56,0.63)	0.34 (0.24,0.43)	0.25	0.60 (SD 0.35) #
Multiple sclerosis	183	0.61 (0.59, 0.62)	0.44 (0.40, 0.49)	0.17	0.55 (0.49, 0.61)*
Parkinson's	147	0.64 (0.62, 0.66)	0.54 (0.49, 0.57)	0.10	0.56 (0.48, 0.63)^

#Edwards 2010 (CCHS 3.1) (102)

*Pohor 2007 (CCHS1.1) (103)

^Jones 2009; Pohor 2009 (CCHS 1.1) (75,78)

Given that utility in individuals with neurological conditions in the general Canadian population are not as low as in the LINC sample, results from the current study's Bland-Altman Plot and regression modeling suggest that discordance between the SF6D and HUI3 would be less in a Canadian population-based neurological sample because the overall HRQoL status in such as sample is not as low as in the LINC study. However, measurement studies in healthier populations still show substantial differences in the distributions of utility scores, regardless of insubstantial differences in mean utility scores. (28)

At five times the frequently cited MID, the mean utility difference of 0.15 (95% CI 0.13, 0.17) was an important discrepancy using any criteria. Among the factors that contributed to the discordance, first and foremost was the difference in utility range.

Although both the SF6D and HUI3 have a theoretical range of 0.00 to 1.00, in reality the scales of the instruments are very different. The HUI3 range is almost double that of the SF6D. Given the very low health status of the sample, the fact that the SF6D is incapable of providing utilities below 0.30 was the leading cause of discordance and an evident limitation of the SF6D. (13,33,34,41,53,61) For example, for the 243 (31.3%) individuals who had HUI3 utility scores below the lower limit of the SF6D (0.30), the mean difference in utility was 0.44 (95% CI 0.42, 0.45). For the remaining 533 individuals with an HUI3 score greater than 0.30; the mean difference in utility was 0.02 (95% CI 0.006, 0.036).

Consistent with the literature, the HUI3 exhibited a mild ceiling effect. (33,34,51,104) The presence of ceiling and/or floor effects is a symptom of reduced variance in the extremities of the utility scale. (49) In the higher health states discordance reversed and the HUI3 provided a higher utility score than the SF6D. The literature frequently speculates that the HUI3's constrained evaluations of the subtler aspects of HRQoL in higher health states is attributable to the omission of domains capturing vitality, impact on social roles, and ability to participate in vigorous and moderate activities; all domains included in the SF6D. (17,41,43,48,61)

A strength of this study was that it went beyond simply describing discordance between the HUI3 and SF6D and investigated the influence of participant characteristics on discordance. This investigation was facilitated by the depth of the data. The *pattern* of discordance between the SF6D and HUI3 was consistent regardless of participant characteristics, type of neurological impairment, or Neuro-QoL domains assessed. Similar to other studies, sociodemographic and general health characteristics were not, generally, substantially associated with discordance; indicating that both instruments were influenced by these factors in similar ways. (61-63) Beyond the discordance associated with HRQoL status, modest differences in discordance were found that aligned with differences in the SF6D and HUI3 descriptive health classification systems. (32,41,53)

HRQoL assessment is constrained by the content of an instrument's descriptive health classification system. For example, the SF6D and HUI3 showed strengths and limitations in HRQoL domains that aligned with the differing "within-the-skin" or "outside-the-skin" emphasis of each instrument. (27,41,43,50) This is consistent with the

existing discordance literature that found the HUI3 to be more sensitive in measuring outcomes of physical and cognitive functioning, whereas the SF6D is more sensitive to outcomes that impact social functioning. (30,34,36) The effect of these differences in sensitivity are seen in the regression models that showed that discordance did not vary by the “within-the-skin” domains of ambulation and cognition; but discordance did vary by “outside-the-skin” domains such as “ability to participate in activities” and “satisfaction in social roles”.

Differences in the descriptive health classification systems were emphasized by the lack of mapping between SF6D and HUI3 domains (Table 4.3). The SF6D and HUI3 instruments contain only a few dimensions defined in loosely similar ways; measuring different components of health with limited and imperfect overlap. (12,27) Additionally, contextual qualifying characteristic such as subjectivity of domain items and differing recall periods, alter how participants consider the extent and duration of impairment. The SF6D evaluates activities within its physical function domain by asking respondents to gauge whether they are limited “a little”, “a lot”, or “not at all”. (26) Whereas the HUI3 items captures health states using statements that leave little room for subjective evaluation. (22) The impact of a condition is an important considerations for many neurological conditions that have inconsistent symptoms from day to day (Parkinson’s disease and multiple sclerosis); are episodic in nature (migraine and epilepsy); and where social functioning and effect on daily life, temper evaluations of impairment. (74,105,106) Even within the SF6D and HUI3’s most “similar” domains of physical function (ambulation/dexterity) and mental (emotional) health there were stark contrasts in the distribution of participants within these domains. For example, in participants in the upper 10% of the utility distribution, only 37% and 38% of individuals reported full functioning in the SF6D’s physical functioning and mental health domains respectively. By comparison the same individuals reported full functioning 99%, 96% and 91% in the HUI3’s ambulation, dexterity and emotion domains (Table 4.8).

In summary, different operationalizations of HRQoL, that include different domains, with different numbers of items and response levels, and different contextual framing, provided participants with different options to describe their HRQoL and contributed to the discordance observed. While these descriptive system differences do not

differ from those found in the evidence, this study confirms that they do indeed contribute to discordance in evaluations of HRQoL in persons with neurological conditions.

Implications: Choice of instrument could influence decision-making

The choice of instrument could influence decision-making. Despite sharing a common purpose, the SF6D and HUI3 vary in many respects in their content and structure, resulting in discordant evaluations of HRQoL. Neither instrument is specifically right or wrong; merely different. (27,30,34,36) The implication of this “difference”, however, is important as healthcare decisions could be affected by the choice of instrument.

The presence of substantial discordance means that SF6D and HUI3 are not interchangeable measures of HRQoL. (27,30,34,36) Depending on the HRQoL status of the sample, the SF6D and HUI3 could provide widely contradictory utility estimates, and *either* instrument could over or underestimate HRQoL status. Additionally, the fact that the instruments showed different associations with neurologically relevant domains implies that either instrument may favour or disadvantage individuals depending on the type of impairment. For instance, if an intervention is designed to increase participation and satisfaction in social roles, the HUI3 may not adequately capture important change in this domain (where change occurred) and may render the intervention ineffective.

In the absence of one instrument being declared the “gold standard”, and given the fact that each instrument captured distinct elements of HRQoL and that sensitivity varied with HRQoL status, consideration of the population and outcomes of interest should be carefully considered when selecting an instrument for a particular context. Further, in heterogeneous populations with low HRQoL status and a broad distribution of utility, the use of more than one utility instrument or the addition of a condition-specific instrument may be required to provide a more comprehensive description of HRQoL.(6,10,11,73,107)

It is impossible to fully disentangle the unique contribution of instrument differences (e.g. range effects, descriptive health classification system, and scoring system) on discordance. In combination, these factors accumulate and interact to produce the substantial systematic variation in the discordance observed with overall HRQoL status; a finding that is likely robust despite the limitations of our study. There are a

number of potential limitations that may affect the results of this study. Two main possible sources of error include: 1) the representativeness of the sample because it is from a non-random sample, and 2) the 20% non-completion rate suggesting a level of survey fatigue. The representativeness of the LINC sample has been compared to the Statistics Canada's Survey of Living with a Neurological Condition in Canada (SLNCC). (92) Generally, participants in the LINC study are slightly older, more educated, and have greater disability than the individuals in the SLNCC. However, in measurement studies representativeness of the sample is less of an issue because selection bias would impact both instruments equally. Therefore within-subject comparisons of utility difference permitted evaluation of discordance without potential confounding effects of inter-subject variations and are unlikely to alter the confidence in the results. (63) Because the order of the instruments was not altered during the administration of the LINC survey, the high rate of non-completion suggests a potential loss of accuracy as individuals progressed through the survey. It is possible that this loss of accuracy could increase measurement error and, subsequently, discordance. However, the strength of the association between discordance and HRQoL is very strong, and therefore, it is unlikely that this limitation would alter the study's conclusions.

The cross-sectional nature of the data precluded the evaluation of the responsiveness of the HUI3 and SF6D to change. The responsiveness of an instrument relates to its ability to detect a change in HRQoL status over time when change *has* occurred. (49) Capturing change over time or change following the implementation of an intervention is valuable. Understanding how responsive the SF6D and HUI3 are in neurological conditions is an important next step in understanding the appropriateness of the use of these instruments in this context.

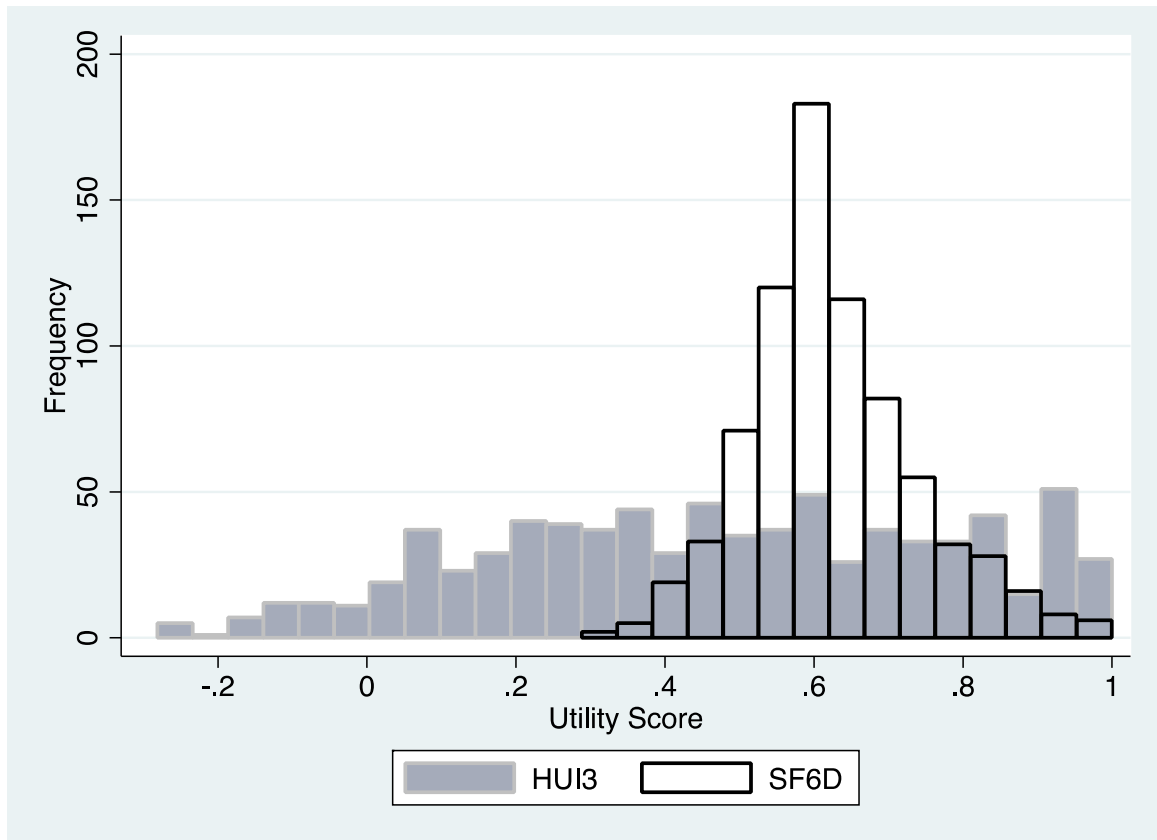
The strong non-linear associations between discordance and overall HRQoL status (mean utility) complicated the evaluation of participant characteristics on variations in discordance. Spline modeling techniques were employed to improve model fit. Three linear splines were chosen for reasons of parsimony. Other knot locations and number of knots were also examined; however, they did not improve model fit nor alter the conclusions in anyway. Additionally, adjusting the base regression model based on

regression diagnostics did influence the regression coefficients marginally but did not impact the interpretation of results or conclusions.

4.5 CONCLUSIONS

Discordance between the SF6D and HUI3 was substantial and varied systematically with the overall HRQoL status. This study adds to the growing body of evidence regarding the methodological differences of preference-based health related quality of life instruments. The clinically important differences found between the SF6D and HUI3 further cast doubt on whether utility estimates between instruments are comparable or universally valid. Consideration of the sample in terms of HRQoL status, population characteristics, and outcome of interest should inform the choice of instrument. The SF6D and HUI3 measure a different array of HRQoL domains and each provides unique information in the evaluation of the impact and the burden of neurological conditions on the everyday lives of Canadians.

Figure 4.1 Distributions of the SF6D and HUI3



* The Shapiro-Wilk test for normality confirms that neither instrument's utility scores is normally distributed ($p < 0.001$)

Figure 4.2 Scatter plot of the SF6D and HUI3 with regression line

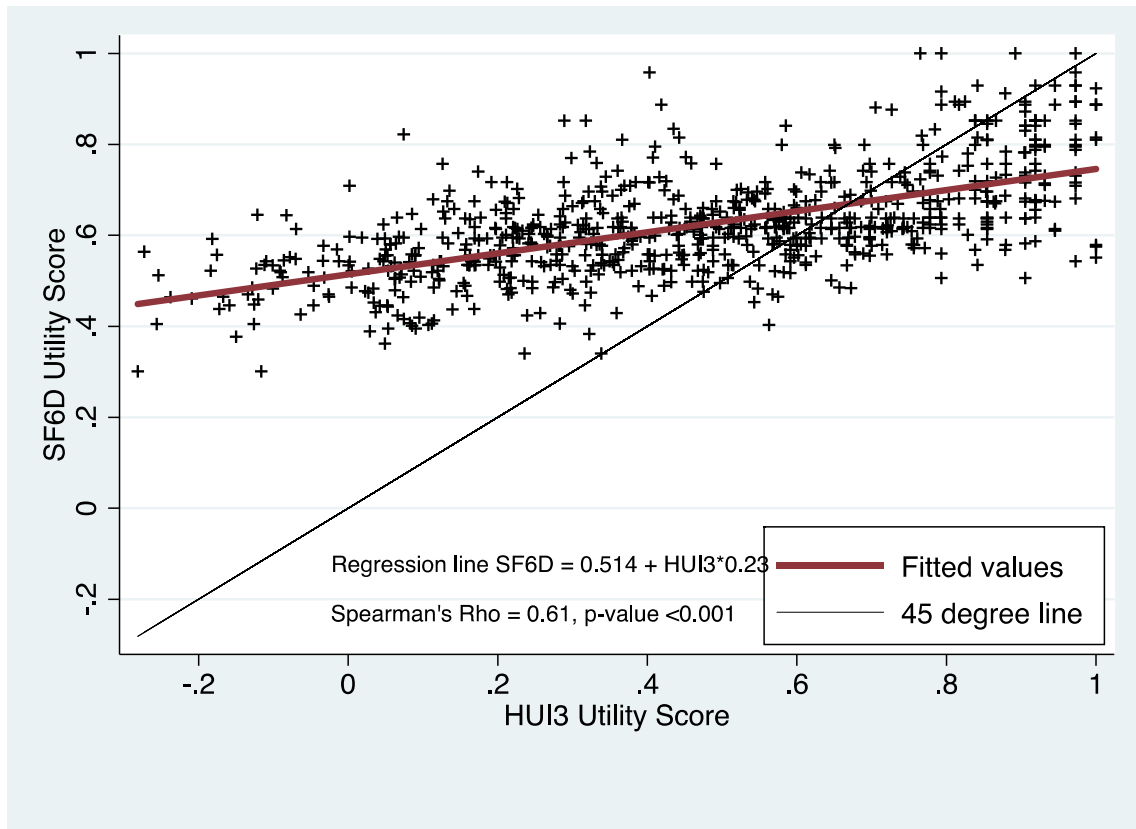
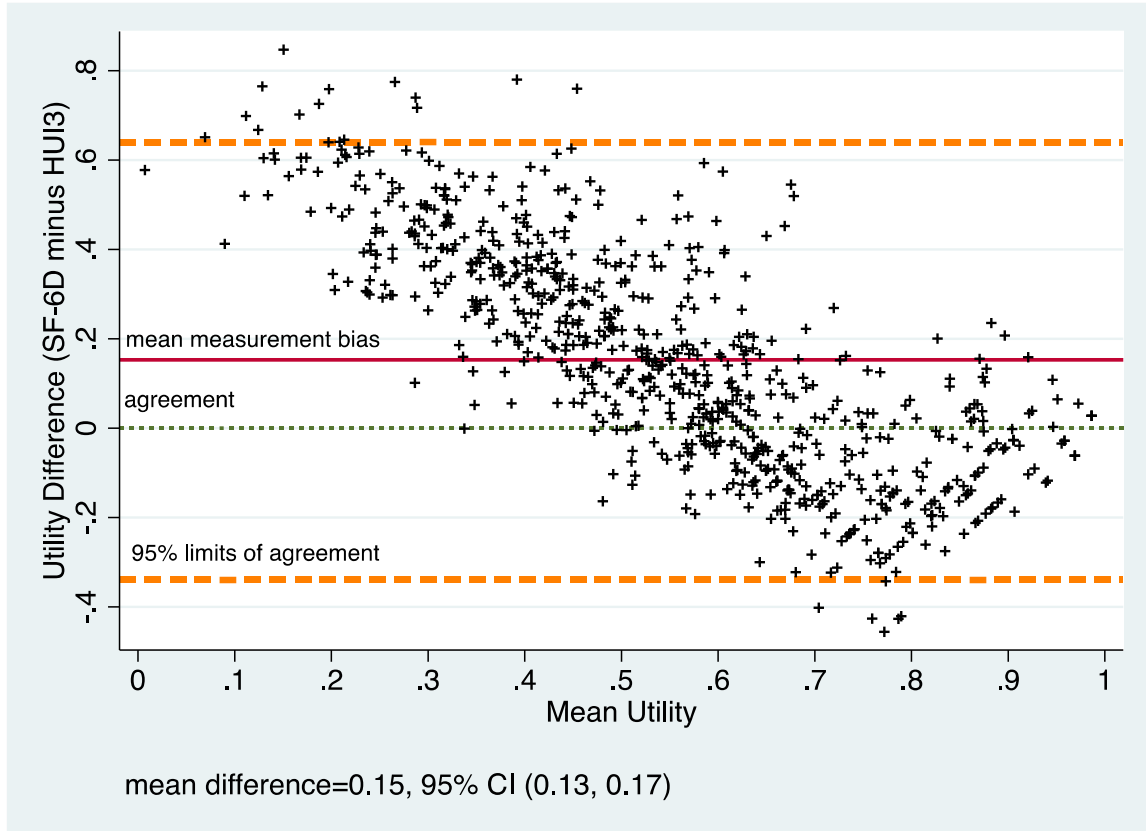


Figure 4.3 Bland-Altman plot of the SF6D and HUI3



Interpretation of Figure:

- The dashed line represents 0 utility difference (agreement). Positive values of dUTY (points above the dashed line of agreement) indicate the HUI3 provides a *lower* utility score than the SF6D, and negative dUTY values indicate the HUI3 provides a *higher* score than the SF6D. Therefore, for individuals between mean utility 0 and 0.6, the HUI3 provides a lower utility score than the SF6D. Difference in utility *decreases* rapidly towards mean utility of 0.6-0.7. Once crossing the zero line, dUTY increases to a mean utility of 0.8 and then declines again. Between dUTY values of 0.65 (approx.) and 1.00 the HUI3 provides a higher score than the SF6D.
- The red line indicates the mean measurement bias and indicates that overall, the HUI3 provides a utility score that is 0.15 lower than the SF6D.
- 95% limits of agreement are mean utility difference ± 1.96 (SD of mean utility difference) and represent the expected limits of agreement for 95% of the observations. These wide limits suggest poor agreement between the SF6D and HUI3.

Table 4.1 Characteristics of the study population

Sample Characteristics (n=766)		n (%)	SF6D (95%CI)	HUI3 (95%CI)	Difference*
Sex	Male	244 (35)	0.62 (0.61, 0.64)	0.43 (0.39, 0.47)	0.19
	Female	444 (65)	0.63 (0.62, 0.64)	0.49 (0.46, 0.52)	0.13
Age	<26	44 (6)	0.66 (0.63, 0.70)	0.49 (0.39, 0.59)	0.18
	27-55	368 (53)	0.61 (0.60, 0.62)	0.47 (0.44, 0.50)	0.14
	56-65	163 (24)	0.62 (0.60, 0.64)	0.44 (0.39, 0.49)	0.18
	>65	115 (17)	0.66 (0.64, 0.68)	0.51 (0.45, 0.56)	0.15
Education	No high school	45 (7)	0.60 (0.56, 0.64)	0.37 (0.27, 0.46)	0.24
	High school	78 (12)	0.61 (0.59, 0.63)	0.41 (0.34, 0.47)	0.20
	Some post sec.	53 (8)	0.62 (0.58, 0.65)	0.41 (0.34, 0.49)	0.21
	Post secondary	491 (73)	0.63 (0.62, 0.64)	0.50 (0.48, 0.53)	0.13
Marital status	Married/common law	429 (62)	0.63 (0.62, 0.64)	0.50 (0.47, 0.53)	0.13
	Single/never married	151 (22)	0.62 (0.60, 0.64)	0.44 (0.39, 0.48)	0.18
	Widow/separated/divorced	107 (16)	0.60 (0.58, 0.62)	0.41 (0.35, 0.47)	0.19
Employment	Working	254 (36)	0.66 (0.64, 0.67)	0.60 (0.56, 0.63)	0.06
	Not working due to health	264 (37)	0.58 (0.57, 0.59)	0.33 (0.30, 0.36)	0.25
	Not working – other	193 (27)	0.65 (0.63, 0.66)	0.51 (0.46, 0.55)	0.14
Self-rated Health	Excellent	44(6)	0.73 (0.68, 0.77)	0.60 (0.50, 0.69)	0.13
	Very good	171(22)	0.68 (0.66, 0.69)	0.63 (0.58, 0.67)	0.05
	Good	304(39)	0.63 (0.62, 0.64)	0.51 (0.48, 0.54)	0.12
	Fair	194(25)	0.58 (0.57, 0.59)	0.35 (0.31, 0.39)	0.23
	Poor	57(7)	0.50 (0.48, 0.53)	0.13 (0.06, 0.19)	0.38
Co-morbid (non-neuro) conditions	0	414(53)	0.65 (0.64, 0.66)	0.53 (0.59, 0.56)	0.12
	1	265(34)	0.60 (0.59, 0.62)	0.43 (0.40, 0.47)	0.17
	2-3	96(13)	0.57 (0.55, 0.59)	0.34 (0.28, 0.40)	0.23
Time since Diagnosis	0-5 years	146 (19)	0.63 (0.47, 0.57)	0.52 (0.47, 0.57)	0.12
	6-10 years	128 (16)	0.60 (0.58, 0.62)	0.43 (0.38, 0.49)	0.16
	> 10 years	502 (65)	0.63 (0.62, 0.64)	0.47 (0.44, 0.50)	0.16
Health in past year	Health same or better	510 (66)	0.65(0.64, 0.66)	0.53 (0.50, 0.55)	0.12
	Health declined	265 (34)	0.57(0.52, 0.58)	0.36 (0.33, 0.40)	0.21

* Difference was compared using two-sided paired t-tests. Mean differences between the SF6D and HUI3 are all significant at a p-value <0.001 except Self-rated Health: excellent p-value 0.003 and Self-rated Health: very good p-value 0.007

† Total non-neurological comorbid conditions: diabetes, heart disease and depression

Table 4.2 Frequency and utility of neurological conditions included in the sample

Neurological conditions ‡ (n = 766)	n	SF6D (95%CI)	HUI3 (95%CI)	Differ[§]
Episodic	236	0.62 (0.60, 0.64)	0.49 (0.44, 0.52)	0.14
Migraine	159	0.60 (0.58, 0.62)	0.44 (0.40, 0.39)	
Epilepsy	117	0.65 (0.62, 0.67)	0.52 (0.46, 0.59)	
Early onset/congenital **	146	0.62 (0.60, 0.63)	0.44 (0.39, 0.48)	0.18
Spina bifida & hydrocephalus	62	0.62 (0.59, 0.65)	0.47 (0.40, 0.54)	
Cerebral Palsy	14	0.58 (0.52, 0.63)	0.44 (0.29, 0.60)	
Tourette's	8	0.63 (0.53, 0.74)	0.50 (0.25, 0.75)	
Sudden onset/traumatic	149	0.60 (0.58, 0.62)	0.38 (0.33, 0.43)	0.22
Brain injury	108	0.61 (0.59,0.63)	0.38 (0.32,0.44)	
Spinal cord injury	47	0.59 (0.56,0.62)	0.37 (0.29, 0.45)	
Stroke	38	0.59 (0.56,0.63)	0.34 (0.24,0.43)	
Degenerative	431	0.62 (0.61, 0.63)	0.46 (0.43, 0.49)	0.16
Multiple sclerosis	183	0.61 (0.59, 0.62)	0.44 (0.40, 0.49)	
Parkinson's	147	0.64 (0.62, 0.66)	0.54 (0.49, 0.57)	
Muscular dystrophy **	65	0.61 (0.59, 0.63)	0.38 (0.31, 0.45)	
ALS	26	0.58 (0.55, 0.61)	0.26 (0.17, 0.36)	
Alzheimer's	11	0.61 (0.54,0.68)	0.31 (0.11, 0.51)	
Huntington's	7	0.61 (0.53, 0.69)	0.58 (0.33, 0.82)	
Other				
Dystonia	43	0.65 (0.61, 0.69)	0.56 (0.47, 0.64)	
Brain/spinal cord tumor	24	0.60 (0.54,0.65)	0.42 (0.29,0.55)	

‡ Neurological condition type categories are not mutually exclusive.

§ Difference was compared using two-sided paired t-tests. Mean differences between the SF6D and HUI3 are all significant at a p-value <0.001

** Muscular dystrophy is also included in congenital/early onset category

Table 4.3 Spearman Correlation Matrix for SF6D and HUI3 HRQoL domains ^a

	SF6D	Physical	Role	Social	Pain	Mental	Vitality
HUI3	0.61	0.50	<i>0.35</i>	0.51	<i>0.42</i>	0.28	0.34
Vision	0.03	0.10	0.05	-0.03	0.06	-0.05	0.04
Hearing	0.14	0.06	0.11	0.13	0.10	0.08	0.12
Speech	0.17	0.20	0.08	0.15	0.11	0.07	0.09
Cognition	0.29	-0.03	0.19	0.33	0.19	0.31	0.19
Ambulation	0.31	0.60	0.22	0.18	0.19	-0.05	0.13
Dexterity	0.28	0.50	0.16	0.18	0.17	-0.04	0.13
Emotion	0.50	0.05	0.24	<i>0.48</i>	0.26	0.50	<i>0.36</i>
Pain	0.56	0.26	0.31	<i>0.37</i>	0.70	0.24	0.27

a. Correlations are between the HUI3 multi-attribute scores and weighted SF6D domain scores

Shaded correlations = correlations between domains considered as “similar”

Bolded correlations indicate strong (>0.50) correlations

Bold & italicized correlations = moderate (>0.35 to 0.50) correlation

Table 4.4 Distribution of participant responses (%) across the SF6D and HUI3 health domain levels^{††}

Table 4.4a Proportion^{‡‡} of participants by HUI3 domain levels (%) (n=776)								
Level	Vision %	Hearing %	Speech %	Cognition %	Ambulation %	Dexterity %	Emotion %	Pain %
1	22.2	87.0	80.9	42.4	45.6	54.3	47.0	20.8
2	71.8	1.6	7.2	6.3	18.9	26.4	34.3	29.5
3	2.1	2.6	10.6	15.0	13.5	4.9	14.8	31.2
4	2.1	6.8	0.5	22.3	9.2	10.6	2.8	15.2
5	1.7	0.6	0.77	13.3	3.6	2.5	1.0	3.6
6	0.3	1.4	n/a	0.8	9.2	1.4	n/a	n/a

Table 4.4b Proportion of participants by SF6D domain levels (%) (n=776)						
Level	Physical Function %	Role limitation %	Social functioning %	Pain %	Mental health %	Vitality %
1	10.1	11.5	21.3	16.2	17.5	2.2
2	15.7	21.4	22.6	13.8	29.5	16.2
3	23.1	6.4	29.9	28.1	32.2	29.3
4	7.4	60.7	19.1	20.9	15.1	31.5
5	31.1	n/a	7.2	16.5	5.7	20.9
6	12.8	n/a	n/a	4.5	n/a	n/a

^{††} Tables represent the proportion of individuals in each level of each SF6D and HUI3 attribute. Levels are based on Likert-type scales where Level 1 indicates no impairment in the health domain and impairment increases through increasing levels within the domain.

^{‡‡} Proportions may not sum to 100% due to rounding

Table 4.5 Multiple regression modeling associations between participant characteristic factors and utility difference (*dUTY*)

		Model 1 ^a			Model 2 ^b			
		n (%)	β Coeff ^{cd}	p-value	F-test	β Coeff ^{cd}	p-value	F-test
Sex	Male	244 (35)	-0.016	0.122	0.122	-0.011	0.266	0.265
	Female	444 (65)	Reference category			Reference category		
Age	<26	44 (6)	0.062	0.003	<0.001	0.062	0.003	<0.001
	27-55	368 (53)	Reference category			Reference category		
	56-65	163 (24)	0.021	0.081		0.021	0.093	
	>65	115 (17)	0.047	0.001		0.044	0.002	
Education	No high school	45 (7)	0.009	0.667	0.483	0.002	0.908	0.512
	High school	78 (12)	0.022	0.157		0.023	0.153	
	Some post sec.	53 (8)	0.015	0.422		0.013	0.500	
	Post secondary	491 (73)	Reference category			Reference category		
Marital status	Married/common law	429 (62)	Reference category		0.625	Reference category		0.618
	Single/never married	151 (22)	0.012	0.332		0.013	0.327	
	Widow/separated/divorced	107 (16)	0.002	0.836		0.003	0.832	
	Working	254 (36)	Reference category		0.004	Reference category		0.254
Employment	Not working due to health	264 (37)	0.012	0.329		0.010	0.398	
	Not working – other	193 (27)	0.041	0.001		0.025	0.100	
	0	414(53)	Reference category		0.050	Reference category		0.015
Co-morbid	1	265(34)	-0.017	0.098		-0.022	0.050	
	2-3	96(13)	-0.034	0.027		-0.044	0.008	
Time since	0-5 years	146 (19)	-0.012	0.318	0.224	-0.021	0.087	0.058
	6-10 years	128 (16)	-0.021	0.112		-0.027	0.042	
Diagnosis	> 10 years	502 (65)	Reference category			Reference category		
	Health same or better	510 (66)	Reference category		<0.001	Reference category		<0.001
Health in past year	Health declined	265 (34)	-0.037	0.000		-0.040	0.000	

- Model 1: $dUTY = \alpha + f(\beta_{mean\ utility}) + \beta_j x_j$, where f is the spline function of 3 linear splines of mean utility, and $\beta_j x_j$ are covariates of interest
- Model 2: $dUTY = \alpha + f(\beta_{mean\ utility}) + \beta_2 sex + \beta_3 age + \beta_j x_j$
- Significance based on a minimally important difference (effect size) of >0.03 and statistical significance p-value <0.05. Coefficient represents utility difference associated with the category of interest versus the reference category.
- Coefficient represents change in utility difference for the characteristic in relation to the reference category

Table 4.6 Multiple regression modeling associations between neurological condition type and utility difference (dUTY)

Neurological condition type ^e	Model 1 ^a		Model 2 ^b		Model 3 ^c		
	No.	β Coeff ^j	p-value	β Coeff ^j	p-value	β Coeff ^j	p-value
Episodic	236	-0.023	0.024	-0.017	0.136	-0.024	0.065
Congenital / early on-set	146	0.006	0.632	0.009	0.486	0.005	0.702
Traumatic / sudden on-set	149	0.008	0.526	0.013	0.320	0.013	0.392
Degenerative	431	-0.006	0.512	-0.012	0.243	-0.012	0.357

a. Model 1: dUTY = $\alpha + f(\beta_{mean\ utility}) + \beta_j X_j$, where f is the spline function of 3 linear splines of mean utility, and $\beta_j X_j$, are covariates of interest

b. Model 2: dUTY = $\alpha + f(\beta_{mean\ utility}) + \beta_2 sex + \beta_3 age + \beta_j X_j$

c. Model 3: dUTY = $\alpha + f(\beta_{mean\ utility}) + \beta_2 sex + \beta_3 age + \beta_4 episodic + \beta_5 congenital \& early\ onset + \beta_6 traumatic \& sudden\ onset + \beta_7 degenerative$

d. Coefficient represents change in utility difference if the individual has the neurological condition type versus not having that type.

e. Neurological condition type categories are not mutually exclusive – study participants can be classified in more than one category

Table 4.7 Multiple regression modeling associations between neurological condition symptoms and impairments (Neuro-QoL scale variables) and utility difference (dUTY)

Neuro-QoL variables	No.	Model 1 ^a			Model 2 ^b			Model 3 ^{c,h}		
		β Coeff ^{d,e}	p-value	β Coeff ^{d,e}	p-value	β Coeff ^{d,e}	p-value	β Coeff ^{d,e}	p-value	
Upper Extremity ^g	763	-0.018	0.000	-0.015	0.007	-0.003	0.633			
Lower Extremity ^g	762	-0.017	0.001	-0.013	0.019	-0.002	0.724			
Cognition – general ^g	753	0.008	0.145	0.008	0.174	-0.022	0.006			
Cognition – executive function ^g	747	0.007	0.229	0.008	0.154	-0.006	0.442			
Fatigue ^f	763	-0.075	0.000	-0.073	0.000	-0.032	0.000			
Sleep disturbance ^f	760	-0.065	0.000	-0.067	0.000	-0.028	0.000			
Depression ^f	746	-0.068	0.000	-0.064	0.000	-0.034	0.001			
Anxiety ^f	749	-0.060	0.000	-0.057	0.000	-0.031	0.000			
Emotional behavioural dyscontrol ^f	748	-0.030	0.000	-0.026	0.000	0.010	0.105			
Positive affect & wellbeing ^g	746	0.055	0.000	0.053	0.000	-0.016	0.078			
Stigma ^f	742	-0.043	0.000	-0.041	0.000	-0.004	0.574			
Ability to participate in social roles ^g	719	0.074	0.000	0.070	0.000	0.036	0.000			
Satisfaction with social roles ^g	719	0.075	0.000	0.072	0.000	0.020	0.081			

a. Model 1: $dUTY = \alpha + f(\beta_{mean\ utility}) + \beta_j x_j$, where f is the spline function of 3 linear splines of mean utility, and $\beta_j x_j$ are covariates of interest

b. Model 2: $dUTY = \alpha + f(\beta_{mean\ utility}) + \beta_2 sex + \beta_3 age + \beta_j x_j$

c. Model 3: Ran one multiple regression model: $dUTY = \alpha + f(\beta_{mean\ utility}) + \beta_2 sex + \beta_3 age + \beta_j x_j$ (all Neuro-QoL included)

d. Regression coefficient represents change in utility difference for 1 standard deviation change in the associated Neuro-QoL variable

e. Significance based on a minimally important difference (effect size) of >0.03 and statistical significance p -value <0.05 .

f. Higher neuro-QoL score relates to *worse* self-reported health

g. Higher neuro-QoL score relates to *better* self-reported health.

h. Collinearity: A number of similar Neuro-QoL variables are collinear with each other, however test of collinearity in Model 3 did not indicate collinearity as an issue in associations with utility difference

NOTE: The regression coefficients predicting utility difference ($dUTY$) are not a continuous variable in the true sense. $dUTY$ measures both the magnitude and the *direction* of discordance between the SFOD and HUI3. The degree of discordance does not simply increase as values of $dUTY$ increases through the range of negative and positive values. Rather, zero represents a “null” value of concordance, and discordance increases in either direction away from zero. For example, a negative coefficient does not indicate a simple inverse relationship between $dUTY$ and a covariate (and a decrease in utility difference). The coefficient represents a change in the intercept. Because the predicted plot of utility difference crosses the zero line (agreement) – a coefficient can be associated with both decrease and increase in utility difference depending on its location along the mean utility axis. Therefore caution should be used if interpreting

Table 4.8 Distribution of participant responses (%) in SF6D and HUI3 health domain levels in individuals with an HUI3 score of 0.90 or greater (n=79) §§

Table 4.8a Proportion*** of participants by HUI3 domain levels (%)								
Level	Vision %	Hearing %	Speech %	Cognition %	Ambulation %	Dexterity %	Emotion %	Pain %
1	27.9	100	100	81.0	98.7	96.2	91.1	67.1
2	72.1†††	0	0	0	1.27	3.8	8.9	32.9‡‡‡
3	0	0	0	19.0§§§	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	n/a	0	0	0	n/a	n/a

Table 4.8b Proportion of participants by SF6D domain levels (%)						
Level	Physical Functioning %	Role limitation %	Social functioning %	Pain %	Mental health %	Vitality %
1	36.7	51.9	53.2	48.1	38.0	5.1
2	39.2	20.3	29.1	22.8	27.9	40.51
3	15.2	16.5	11.4	17.7	24.0	39.2
4	0	11.4	2.5	6.3	7.6	13.9
5	5.1	n/a	3.8	3.8	2.5	1.27
6	3.8	n/a	n/a	1.3	n/a	n/a

§§ Tables represent the proportion of individuals in each level of each SF6D and HUI3 attribute. Levels are based on Likert-type scales where Level 1 indicates no impairment in the health domain and impairment increases through increasing levels within the domain.

*** Proportions may not sum to 100% due to rounding

††† Can see newspaper or friends on other side of street with help of glasses

‡‡‡ Mild to moderate pain that prevents no activities

§§§ Somewhat forgetful, but able to think clearly and solve day to day problems

CHAPTER 5. CONCLUSION

Preference-based health-related quality of life instruments are widely used for evaluating treatments and interventions across a wide variety of health conditions. (12,27) Alternative measures share a similar theoretical rationale and are designed to measure the same underlying concept. Nevertheless, they can and do differ in important respects. Utility instruments are based on different models of health; they include a different number and variety of HRQoL domains; and use different preference elicitation methods and scoring algorithms. Differences in the way utility instruments define, describe and value HRQoL affects the validity, interpretability, and meaningfulness of utilities generated by these instruments. (8,12,27)

This study examined the performance of two preference-based health related quality of life instruments, the SF6D and HUI3, in Canadians with neurological conditions. Although agreement between these instruments has been studied before, it has not been assessed in a diverse sample of neurological conditions. The degree to which the SF6D and HUI3 are sensitive to variation in domains of HRQoL, relevant for persons with neurological conditions, was assessed and the extent of agreement / discordance evaluated.

The study found that SF6D and HUI3 did not evaluate HRQoL in persons with neurological conditions in the same way. Substantial and clinically important differences were seen in utility scores, distributions, and utility range. A key finding was that discordance between the SF6D and HUI3 varied systematically with overall HRQoL status. Both instruments were sensitive to HRQoL domains relevant to persons with neurological conditions. However, neither instrument evaluated a full complement of HRQoL domains that fully describe the impact of neurological conditions. The evaluations of HRQoL provided by the SF6D and HUI3, therefore, cannot be regarded as fully comprehensive. (8) Further, study findings suggested that either instrument may *favour or disadvantage* individuals depending on an individual's HRQoL status. Generally, in individuals with low utility, the HUI3 will provide a lower utility score than the SF6D. In individuals with higher utility, the reverse is true.

There is no consensus, even within a single condition, on a common HRQoL measure. (31,35,37) Therefore this study also sought to inform the choice of instrument. Because each instrument captures distinct elements of HRQoL and discordance varied with HRQoL status, the HRQoL status of a sample, participant characteristics and the outcomes of interest should all guide the choice of instrument. The HUI3 was more sensitive to domains of physical and cognitive functioning, whereas, the SF6D had greater association with domains that impact social functioning and vitality. Therefore, consideration of the presence of these types of characteristics in a study sample will assist in choosing the most relevant and valid instrument. Additionally, the HUI3 described disutility in greater detail for individuals with low health states (below utilities of approximately 0.65) and also provided greater magnitude of differences between degrees (severities) of HRQoL impairment. The HUI3 would therefore benefit studies where there is a need to capture subtle changes in HRQoL in populations with lower HRQoL status; especially decrements in HRQoL. The SF6D was able to better differentiate health states with a greater spread of values above utilities of 0.65 and is therefore suitable for populations with higher HRQoL status. The choice of which instrument to use should, as a result, be in careful consideration of the strengths and limitation of each instrument in a particular context.

Researchers, program evaluators, policy makers, and clinicians will benefit from better understanding the context within which the utility instruments agree and diverge when describing HRQoL of individuals with multiple, complex health concerns such as neurological conditions. Those with an interest in treating, surveillance, modeling, simulating, and planning for the future care of Canadians with neurological conditions will, it is hoped, benefit from this study's detailed evaluation of the SF6D and HUI3.

5.1 LIMITATIONS

The study data had a number of limitations. First, the data was from a non-random, voluntary sample and selection bias was possible. Based on comparisons of the LINC study to the Survey of Living with a Neurological Condition in Canada (SLNCC), evidence suggests that the LINC sample is not fully representative of Canadians with neurological conditions. Among the differences, is that the LINC sample was slightly

older, more highly educated, and had greater disability. (92) However, in measurement studies representativeness of the sample is less of an issue because selection bias would impact both instruments equally. Therefore within-subject comparisons of utility difference permitted evaluation of discordance without potential confounding effects of inter-subject variations and are unlikely to alter the confidence in the results, as any potential biases would involve both instruments in a similar way. (63) Second, the diagnosis and morbidity information is self-reported, leading to a potential for misclassification. Similar to representativeness, this would impact both measures in a similar way and is not likely to alter conclusions.

Third, the order of individual instruments and scales in the LINC survey was not random and, in light of the 20% non-completion rate, there is a risk of survey fatigue and potential loss of accuracy as individuals progressed through the survey. It is possible that this loss of accuracy could increase measurement error and, subsequently, discordance. However, the strength of the association between discordance and HRQoL is very strong, and therefore, it is unlikely that this limitation would alter the study's conclusions.

Fourth, the cross-sectional nature of the data precluded evaluation of discordance in relation to responsiveness, which is important in evaluating change in HRQoL. Lastly, the impacts of different preference elicitation methods and scoring functions on differential measurement were suspected, as a result of this study, but were beyond the scope of the study and data to describe or quantify.

5.2 IMPLICATIONS FOR FUTURE RESEARCH

Two preference-based utility instruments, the SF6D and HUI3, were compared, as a comparison of opportunity, not design. It would be beneficial to assess the performance of other utility instruments, such as the EQ-5D, to compare utility estimates across a broader range of instruments. Evaluating the responsiveness of the SF6D and HUI3 in a longitudinal study of individuals with neurological conditions, would further the understanding of how these instruments perform, agree, and evaluate changes in health status and treatment effects over time. Because social functioning, role limitations and vitality were important in describing the HRQoL of the participants in this study, further

research is needed to evaluate the sensitivity and responsiveness of utility instruments specifically in capturing changes in these important HRQoL domains.

In summary

The ability of utility instruments to simultaneously summarize the net gain (or loss) in HRQoL into a single score is beneficial in neurological conditions because these conditions, and their treatments, often impact multiple domains. However, lack of agreement in, and omissions of important HRQoL domains significantly limit the usefulness of a single instrument to comprehensively evaluate HRQoL in persons with neurological conditions. The inclusion of a condition-specific instrument, in conjunction with utility instruments, will therefore enrich the evaluation of HRQoL in complex, heterogeneous populations.

REFERENCES

1. The burden of neurological disease, disorders and injuries in Canada. Canadian institute for health information : Available from: http://www.cpa.ca/cpasite/userfiles/Documents/Practice_Page/Burden_neuro_diseases_en.pdf. Accessed 20 September 2012.
2. A brain strategy for Canada. Neurological health charities Canada : Available from: http://mssociety.ca/en/pdf/socact_BrainStrategy.pdf. Accessed 1 February 2013.
3. Karlsen H, Tandenberg E, Arslan D, Larsen P. Health related quality of life in Parkinson's disease: A prospective longitudinal study. *Journal of Neurology, Neurosurgery & Psychiatry* 2000, Nov 1;69(5):584-9.
4. Bode RK, Heinemann AW, Butt Z, Stallings J, Taylor C, Rowe M, Roth EJ. Development and validation of participation and positive psychologic function measures for stroke survivors. *Archives of Physical Medicine and Rehabilitation* 2010, Sep;91(9):1347-56.
5. Vossius C, Nilsen OB, Larsen JP. Health state values during the first year of drug treatment in early-stage Parkinson's disease: A prospective, population-based, cohort study. *Drugs Aging* 2009;26(11):973-80.
6. Heffernan C, Jenkinson C. Review paper: Measuring outcomes for neurological disorders: A review of disease-specific health status instruments for three degenerative neurological conditions. *Chronic Illness* 2005, Jun 1;1(2):131-42.
7. Mayo NE, Moriello C, Asano M, Spuy S, Finch L. The extent to which common health-related quality of life indices capture constructs beyond symptoms and function. *Quality of Life Research* 2011, Jun;20(5):621-7.
8. Kuspinar A, Mayo NE. Do generic utility measures capture what is important to the quality of life of people with multiple sclerosis? *Health Qual Life Outcomes* 2013;11(1):71.
9. Carod-Artal FJ. Determining quality of life in stroke survivors. *Expert Rev Pharmacoecon Outcomes Res* 2012, Apr;12(2):199-211.
10. Cella D, Lai JS, Nowinski CJ, Victorson D, Peterman A, Miller D, et al. Neuro-QOL: Brief measures of health-related quality of life for clinical research in neurology. *Neurology* 2012, Jun 5;78(23):1860-7.
11. Cella D, Nowinski C, Peterman A, Victorson D, Miller D, Lai J-S, Moy C. The neurology quality-of-life measurement initiative. *Archives of Physical Medicine and Rehabilitation* 2011, Oct;92(10):S28-36.

12. Fryback, DG. Measuring health-related quality of life; Workshop on advancing social science theory: The importance of common metrics. The national academies, division of behavioral and social sciences and education washington, DC. 2010.
13. Kopec J. A comparative review of four preference-weighted measures of health-related quality of life. *Journal of Clinical Epidemiology* 2003, Apr;56(4):317-25.
14. Guyatt H, Feeny H, Patrick L. Measuring health-related quality of life. *Annals of Intern Medicine* 1993;118:622-9.
15. Ravicki D, Kaplan M. Relationship between psychometric and utility-based approaches to measurement of health-related quality of life. *Quality of Life Research* 1993;2(6):477-87.
16. Lipscomb J, Drummond M, Fryback D, Gold M, Revicki D. Retaining, and enhancing, the QALY. *Value in Health* 2009, Mar;12:S18-26.
17. McDonough CM, Tosteson AN. Measuring preferences for cost-utility analysis: How choice of method may influence decision-making. *Pharmacoeconomics* 2007;25(2):93-106.
18. Conner-Spady B, Suarez-Almazor ME. Variation in the estimation of quality-adjusted life-years by different preference-based instruments. *Med Care* 2003, Jul;41(7):791-801.
19. Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: Learning from psychometrics. *Health Econ* 1999, Feb;8(1):41-51.
20. Lenert L, Kaplan RM. Validity and interpretation of preference-based measures of health-related quality of life. *Med Care* 2000, Sep;38(9 Suppl):II138-50.
21. Torrance GW. Measurement of health state utilities for economic appraisal. *Journal of Health Economics* 1986, Mar;5(1):1-30.
22. Horsman J, Furlong W, Feeny D, Torrance G. The health utilities index (HUI): Concepts, measurement properties and applications. *Health Qual Life Outcomes* 2003;1:54.
23. Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Med Care* 2002, Feb;40(2):113-28.
24. Cheung K, Oemar M, Oppe M, Rabin R. EQ-5D user guide. Version 2.0. : Available from: www.euroqol.org/fileadmin/user.../User_Guide_v2_March_2009.pdf. Accessed 18 January 2013.

25. Brooks R, Rabin R, de Charro F. The measurement and valuation of health status using EQ-5D: A european perspective. Dordrecht, The Netherlands: Kluwer Academic Publishers; 2003.
26. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002, Mar;21(2):271-92.
27. Richardson J, McKie J, Bariola E. Review and critique of health related multi attribute utility instruments. Monash University, Business and Economics, Centre for Health Economics; 2011.
28. Fryback DG, Dunham NC, Palta M, Hanmer J, Buechner J, Cherepanov D, et al. US norms for six generic health-related quality-of-life indexes from the national health measurement study. *Medical Care* 2007, Dec;45(12):1162-70.
29. Kaplan RM, Tally S, Hays RD, Feeny D, Ganiats TG, Palta M, Fryback DG. Five preference-based indexes in cataract and heart failure patients were not equally responsive to change. *Journal of Clinical Epidemiology* 2011, May;64(5):497-506.
30. Pickard AS, Johnson JA, Feeny DH. Responsiveness of generic health-related quality of life measures in stroke. *Quality of Life Research* 2005, Feb;14(1):207-19.
31. Marra A, Marion A, Guh P, Najafzadeh M, Wolfe F, Esdaile M, et al. Not all "quality-adjusted life years" are equal. *Journal of Clinical Epidemiology* 2007, Jun;60(6):616-24.
32. O'Brien BJ, Spath M, Blackhouse G, Severens L, Dorian P, Brazier J. A view from the bridge: Agreement between the SF-6D utility algorithm and the health utilities index. *Health Economics* 2003, Nov;12(11):975-81.
33. Moock J, Kohlmann T. Comparing preference-based quality-of-life measures: Results from rehabilitation patients with musculoskeletal, cardiovascular, or psychosomatic disorders. *Quality of Life Research* 2008, Apr;17(3):485-95.
34. Fisk D. A comparison of health utility measures for the evaluation of multiple sclerosis treatments. *Journal of Neurology, Neurosurgery & Psychiatry* 2005, Jan 1;76(1):58-63.
35. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics* 2004, Sep;13(9):873-84.
36. Langfitt T, Vickrey G, McDermott P, Messing S, Berg T, Spencer S, et al. Validity and responsiveness of generic preference-based HRQOL instruments in chronic epilepsy. *Quality of Life Research* 2006, Jun;15(5):899-914.

37. Hawthorne G, Richardson J, Day NA. A comparison of the assessment of quality of life (aqol) with four other generic utility instruments. *Annals of Medicine* 2001, Jul;33(5):358-70.
38. Rashidi AA, Anis AH, Marra CA. Do visual analogue scale (VAS) derived standard gamble (SG) utilities agree with health utilities index utilities? A comparison of patient and community preferences for health status in rheumatoid arthritis patients. *Health Qual Life Outcomes* 2006;4(1):25.
39. Hays RD, Hahn H, Marshall G. Use of the SF-36 and other health-related quality of life measures to assess persons with disabilities. *Archives of Physical Medicine and Rehabilitation* 2002, Dec;83:S4-9.
40. McPherson K, Myers J, Taylor WJ, McNaughton HK, Weatherall M. Self-valuation and societal valuations of health state differ with disease severity in chronic and disabling conditions. *Med Care* 2004, Nov;42(11):1143-51.
41. Davison SN, Jhangri GS, Feeny DH. Comparing the health utilities index mark 3 (HUI3) with the short form-36 preference-based SF-6D in chronic kidney disease. *Value in Health* 2009, Mar;12(2):340-5.
42. Furlong WJ, Feeny DH, Torrance GW, Barr RD. The health utilities index (HUI) system for assessing health-related quality of life in clinical studies. *Annals of Medicine* 2001, Jul;33(5):375-84.
43. Richardson J. Psychometric validity and multi attribute utility (MAU) instruments. Monash University, Business and Economics; 2010.
44. Whitehurst DG, Bryan S. Another study showing that two preference-based measures of health-related quality of life (EQ-5D and SF-6D) are not interchangeable. But why should we expect them to be? *Value in Health* 2011, Jun;14(4):531-8.
45. Ware E, Donald Sherbourne C. The MOS 36-item short-form health survey (SF-36) conceptual framework and item selection. *Medical Care* 1992;30(6):473-83.
46. Victorson D, Cavazos JE, Holmes GL, Reder AT, Wojna V, Nowinski C, et al. Validity of the neurology quality-of-life (neuro-qol) measurement system in adult epilepsy. *Epilepsy & Behavior* 2014, Feb;31:77-84.
47. SF-6D. SF-36v2 user's manual : Available from: <http://www.qualitymetric.com/Portals/0/Uploads/Documents/Public/SF-6D.pdf>. Accessed 20 September 2012.
48. Ferreira PL, Ferreira LN, Pereira LN. How consistent are health utility values? *Quality of Life Research* 2008, Sep;17(7):1031-42.

49. Seymour J, McNamee P, Scott A, Tinelli M. Shedding new light onto the ceiling and floor? A quantile regression approach to compare EQ-5D and SF-6D responses. *Health Economics* 2009;n/a.
50. Hatoum HT, Brazier JE, Akhras KS. Comparison of the HUI3 with the SF-36 preference based SF-6D in a clinical trial setting. *Value Health* 2004;7(5):602-9.
51. Fryback G, Palta M, Cherepanov D, Bolt D, Kim -S. Comparison of 5 health-related quality-of-life indexes using item response theory analysis. *Medical Decision Making* 2010, Jan 1;30(1):5-15.
52. Marra CA, Woolcott JC, Kopec JA, Shojania K, Offer R, Brazier JE, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the raqol and the HAQ) in rheumatoid arthritis. *Social Science & Medicine* 2005, Apr;60(7):1571-82.
53. McDonough CM, Grove MR, Tosteson TD, Lurie JD, Hilibrand AS, Tosteson ANA. Comparison of EQ-5D, HUI, and sf-36-derived societal health state values among spine patient outcomes research trial (SPORT) participants. *Quality of Life Research* 2005, Jun;14(5):1321-32.
54. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput Biol Med* 1990;20(5):337-40.
55. Feeny D, Wu L, Eng K. Comparing short form 6D, standard gamble, and health utilities index mark 2 and mark 3 utility scores: Results from total hip arthroplasty patients. *Quality of Life Research* 2004, Dec;13(10):1659-70.
56. Raisch DW, Feeney P, Goff DC, Narayan M, O'Connor PJ, Zhang P, et al. Baseline comparison of three health utility measures and the feeling thermometer among participants in the action to control cardiovascular risk in diabetes trial. *Cardiovascular Diabetology* 2012;11(1):35.
57. Kontodimopoulos N, Argiriou M, Theakos N, Niakas D. The impact of disease severity on EQ-5D and SF-6D utility discrepancies in chronic heart failure. *The European Journal of Health Economics* 2011, Aug;12(4):383-91.
58. Obradovic M, Lal A, Liedgens H. Validity and responsiveness of euroqol-5 dimension (EQ-5D) versus short form-6 dimension (SF-6D) questionnaire in chronic pain. *Health Qual Life Outcomes* 2013;11(1):110.
59. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 1999;8(2):135-60.
60. Bland JM, Martin, Altman G. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* 1986, Feb 8:307-10.

61. Quercioli C, Messina G, Barbini E, Carriero G, Fani M, Nante N. Importance of sociodemographic and morbidity aspects in measuring health-related quality of life: Performances of three tools. *The European Journal of Health Economics* 2009, Oct;10(4):389-97.
62. Sogaard R, Christensen FB, Videbæk TS, Bünger C, Christiansen T. Interchangeability of the EQ-5D and the SF-6D in long-lasting low back pain. *Value in Health* 2009, Jun;12(4):606-12.
63. Wee H-L, Machin D, Loke W-C, Li S-C, Cheung Y-B, Luo N, et al. Assessing differences in utility scores: A comparison of four widely used preference-based instruments. *Value in Health* 2007, Jul;10(4):256-65.
64. Robert A, Cherepanov D, Palta M, Dunham C, Feeny D, Fryback G. Socioeconomic status and age variations in health-related quality of life: Results from the national health measurement study. *Journals of Gerontology Series B- Psychological Sciences and Social Sciences* 2009, May 1;64B(3):378-89.
65. Kopec JA, Williams JI, To T, Austin PC. Cross-cultural comparisons of health status in Canada using the Health Utilities Index. *Ethn Health* 2001, Feb;6(1):41-50.
66. Elliott JO, Mares AS. Gender differences in quality of life among canadian adults with epilepsy. *Epilepsy Research* 2012, Jun;100(1-2):42-8.
67. Nowinski CJ, Victorson D, Cavazos JE, Gershon R, Cella D. Neuro-QOL and the NIH toolbox: Implications for epilepsy. *Therapy* 2010, Sep 1;7(5):533-40.
68. Poissant L, Mayo NE, Wood-Dauphinee S, Clarke AE. The development and preliminary validation of a preference-based stroke index (PBSI). *Health Qual Life Outcomes* 2003;1:43.
69. Coons SJ, Rao S, Keininger DL, Hays RD. A comparative review of generic quality-of-life instruments. *Pharmacoeconomics* 2000, Jan;17(1):13-35.
70. Duncan W, Wallace D, Lai M, Johnson D, Embretson S, Laster J. The stroke impact scale version 2.0 : Evaluation of reliability, validity, and sensitivity to change. *Stroke* 1999, Oct 1;30(10):2131-40.
71. Hobart J, Lampin D, Fitzpatrick R, Riazi A, Thompson A. The multiple sclerosis impact scale (MSIS-29): A new patient-based outcome measure. *Brain* 2001;124:962-73.
72. Devinsky O, Vickrey BG, Cramer J, Perrine K, Hermann B, Meador K, Hays RD. Development of the quality of life in epilepsy inventory. *Epilepsia* 1995, Nov;36(11):1089-104.

73. Dodel R, Jönsson B, Reese JP, Winter Y, Martinez-Martin P, Holloway R, et al. Measurement of costs and scales for outcome evaluation in health economic studies of Parkinson's disease. *Movement Disorders* 2014, Feb;29(2):169-76.
74. Brown JS, Neumann PJ, Papadopoulos G, Ruoff G, Diamond M, Menzin J. Migraine frequency and health utilities: Findings from a multisite survey. *Value in Health* 2008, Mar;11(2):315-21.
75. Jones CA, Pohar SL, Warren S, Turpin KV, Warren KG. The burden of multiple sclerosis: A community health survey. *Health Qual Life Outcomes* 2008;6(1):1.
76. Jones CA, Pohar SL, Patten SB. Major depression and health-related quality of life in parkinson's disease. *General Hospital Psychiatry* 2009, Jul;31(4):334-40.
77. Neumann J, Sandberg A, Araki S, Kuntz M, Feeny D, Weinstein C. A comparison of HU12 and HU13 utility scores in alzheimer's disease. *Medical Decision Making* 2000, Oct 1;20(4):413-22.
78. Pohar SL, Allyson Jones C. The burden of parkinson disease (PD) and concomitant comorbidities. *Archives of Gerontology and Geriatrics* 2009, Sep;49(2):317-21.
79. Siderowf A, Ravina B, Glick HA. Preference-based quality-of-life in patients with Parkinson's disease. *Neurology* 2002, Jul 9;59(1):103-8.
80. Versteegh MM, Leunis A, Uyl-de Groot CA, Stolk EA. Condition-Specific preference-based measures: Benefit or burden? *Value in Health* 2012, May;15(3):504-13.
81. Versnel J, Packer T, Weeks LE, Brown J, Godwin M, Hutchinson S, et al. The everyday experience of living with and managing a neurological condition (the LINC study): Study design. *BMC Neurol* 2013;13:30.
82. STATA [computer program]. College Station, Texas: STATA Corp; 2012.
83. Opinio [computer program]. Oslo, Norway: Object Planet Inc.
84. Excel [computer program]. Ontario, Canada: Microsoft Canada.
85. SAS [computer program]. Cary, NC USA: SAS Institute Inc.
86. McCabe C, Brazier J, Gilks P, Tsuchiya A, Roberts J, O'Hagan A, Stevens K. Using rank data to estimate health state utility models. *Journal of Health Economics* 2006, May;25(3):418-31.
87. Kharroubi SA, Brazier JE, Roberts J, O'Hagan A. Modelling SF-6D health state preference data using a nonparametric bayesian method. *Journal of Health Economics* 2007, May;26(3):597-612.

88. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care* 2004, Sep;42(9):851-9.
89. User manual for the quality of life in neurological disorders (neuro-qol) measures, version 1.0. National institute of neurological disorders and stroke (NINDS): Available from:
<http://www.neuroqol.org/Resources/Resources%20documents/Neuro-QOL-UserManual-2013.pdf>. Accessed 1 April 2013.
90. Quatrano LA, Cruz TH. Future of outcomes measurement: Impact on research in medical rehabilitation and neurologic populations. *Archives of Physical Medicine and Rehabilitation* 2011, Oct;92(10):S7-S11.
91. Naess H, Lunde L, Brogger J. The effects of fatigue, pain, and depression on quality of life in ischemic stroke patients: The bergen stroke study. *Vascular Health and Risk Management* 2012, Jun:407.
92. Survey of living with a neurological condition in canada (SLNCC). Statistics Canada : Available from:
<http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5182>. Accessed February 2014.
93. Brown CA, Cheng EM, Hays RD, Vassar SD, Vickrey BG. SF-36 includes less parkinson disease (PD)-targeted content but is more responsive to change than two pd-targeted health-related quality of life measures. *Quality of Life Research* 2009, Nov;18(9):1219-37.
94. Kavirajan H, Hays RD, Vassar S, Vickrey BG. Responsiveness and construct validity of the health utilities index in patients with dementia. *Medical Care* 2009, Jun;47(6):651-61.
95. Lee BB, King MT, Simpson JM, Haran MJ, Stockler MR, Marial O, Salkeld G. Validity, responsiveness, and minimal important difference for the SF-6D health utility scale in a spinal cord injured population. *Value in Health* 2008, Jul;11(4):680-8.
96. Maddigan SL, Feeny DH, Majumdar SR, Farris KB, Johnson JA. Health utilities index mark 3 demonstrated construct validity in a population-based sample with type 2 diabetes. *Journal of Clinical Epidemiology* 2006, May;59(5):472-7.
97. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Quality of Life Research* 2005, Aug;14(6):1523-32.
98. Feeny D, Spritzer K, Hays D, Liu H, Ganiats G, Kaplan M, et al. Agreement about identifying patients who change over time: Cautionary results in cataract and heart failure patients. *Medical Decision Making* 2012, Mar 1;32(2):273-86.

99. Fayers P, Machin D. Quality of life: The assessment, analysis and interpretation of patient-reported outcomes. 2nd Edition ed. Chichester, West Susses, UK: John Wiley & Sons, Ltd; 2007.
100. Shrout E, Fleiss L. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* 1979;86(2):420-8.
101. Leffondre K, Jager J, Boucquemont J, Stel S, Heinze G. Representation of exposures in regression analysis and interpretation of regression coefficients: Basic concepts and pitfalls. *Nephrology Dialysis Transplantation* 2013, Dec 22.
102. Edwards D, Koehoorn M, Boyd A, Levy R. Is health-related quality of life improving after stroke?: A comparison of health utilities indices among canadians with stroke between 1996 and 2005. *Stroke* 2010, May 1;41(5):996-1000.
103. Pohar SL, Jones CA, Warren S, Turpin KV, Warren K. Health status and health care utilization of multiple sclerosis in Canada. *Can J Neurol Sci* 2007, May;34(2):167-74.
104. Busija L, Pausenberger E, Haines TP, Haymes S, Buchbinder R, Osborne RH. Adult measures of general health and health-related quality of life: Medical outcomes study short form 36-item (SF-36) and short form 12-item (SF-12) health surveys, nottingham health profile (NHP), sickness impact profile (SIP), medical outcomes study sh. *Arthritis Care & Research* 2011, Nov;63(S11):S383-412.
105. King-Kallimanis BL, Oort FJ, Nolte S, Schwartz CE, Sprangers MAG. Using structural equation modeling to detect response shift in performance and health-related quality of life scores of multiple sclerosis patients. *Quality of Life Research* 2011, Dec;20(10):1527-40.
106. Fernandez O, Baumstarck-Barrau K, Simeoni -C, Auquier P, on behalf of the MusiQoL study group. Patient characteristics and determinants of quality of life in an international population with multiple sclerosis: Assessment using the musiqol and SF-36 questionnaires. *Multiple Sclerosis Journal* 2011, Oct 1;17(10):1238-49.
107. Guyatt G, Jaeschke R. Reassessing quality-of-life instruments in the evaluation of new drugs. *Pharmacoeconomics* 1997, Dec;12(6):621-6.
108. Ware JE. SF-36.Org; Available from: <http://www.sf-36.org/tools/sf36.shtml>. Accessed 20 September 2012.
109. Whitehurst DGT, Bryan S, Lewis M. Systematic review and empirical comparison of contemporaneous EQ-5D and SF-6D group mean scores. *Medical Decision Making* 2011, Nov 1;31(6):E34-44.

APPENDIX A: Health Utility Index - Mark 3 (HUI3)

The HUI[®] is a generic, indirect preference-based measurement system that provides single and multi-attribute utility index scores that reflect community preferences for the respondent's assessment of his or her health. HUI3 describes 8 dimensions of health status: vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain. Each question has 5-6 levels, which can describe a total of 972,000 unique health states. The preference scoring is based on standard gamble utilities. Following administration of the questionnaire, the results are classified and attribute levels as it relates to health status and the utility formula (based on information collected from the general population) are then converted into single attribute utilities or multi attribute utilities. The range of summary scores for HUI3 is -0.36 to 1.0; where 0 represents death and 1.0 perfect health. (23) The HUI3 scoring system is based on visual analogue scales (VAS) and standard gamble (SG) scores obtained from a random sample of the general population in Hamilton, Ontario (n=504). Subjects were first asked to rate health states on a 100-point visual analogue scale (feeling thermometer) with perfect health at the top and the least desirable state at the bottom. Subjects were then asked to assess a subset of the health states by using the standard gamble valuation technique (chance board) questions. The lottery for the SG consisted of the all-worst HUI3 health state and all best HUI3 health states. (55) To generate the interval scale data for fitting HUI utility functions, the VAS scores were converted to standard gamble valuation, forecast from the VAS by using a statistical power transformation. The basis of the HUI3 is a multiplicative form of utility function (invoked simplifying assumptions from the multi-attribute utility theory), which assumes the same interaction between all attributes and all attribute levels. (32) The HUI is widely used and has been validated across multiple large population surveys. There are currently sixteen versions of the HUI available. Versions differ dependent on the combinations of mode of administration, assessment perspective, duration of assessment period for a given health status, and language. Studies support the HUI's responsiveness, reliability and validity (face, content, construct, convergent, discriminative and predictive) across various populations and contexts. (22)

HUI3 Multi-Attribute Utility Function* on Dead-Healthy Scale

Vision x1 b1	Hearing x2 b2	Speech x3 b3	Ambulation x4 b4	Dexterity x5 b5	Emotion x6 b6	Cognition x7 b7	Pain x8 b8
-----------------	------------------	-----------------	---------------------	--------------------	------------------	--------------------	---------------

*Furlong et al. CEHPA WP#98-11, Table 11, page 76 and Appendix B, Table 1, page 96. (42)

Where x_n is the attribute level and b_n is the attribute utility score. The HUI3 is calculated using the multi-attribute scoring function based on HUI3 individual attributes using the formula provided by the Health Utilities Incorporated that is represented by:

$$(\text{Dead} - \text{Perfect Health}) u^* = 1.371 (b_1 x_n * b_2 x_n * b_3 x_n * b_4 * x_n * b_5 x_n * b_6 x_n * b_7 x_n * b_8 * x_n) - 0.371$$

Where u^* is the utility score, x_n is the attribute level and $b_{1...8}$ is the attribute utility score.

Health Utilities Index – Mark 3

Multi-attribute Health Classification System

Attribute	Level	Description
VISION	1	Able to see well enough to read ordinary newsprint and recognize a friend on the other side of the street, without glasses or contact lenses.
	2	Able to see well enough to read ordinary newsprint and recognize a friend on the other side of the street, but with glasses.
	3	Able to read ordinary newsprint with or without glasses but unable to recognize a friend on the other side of the street, even with glasses.
	4	Able to recognize a friend on the other side of the street with or without glasses but unable to read ordinary newsprint, even with glasses.
	5	Unable to read ordinary newsprint and unable to recognize a friend on the other side of the street, even with glasses.
	6	Unable to see at all.
HEARING	1	Able to hear what is said in a group conversation with at least three other people, without a hearing aid.
	2	Able to hear what is said in a conversation with one other person in a quiet room without a hearing aid, but requires a hearing aid to hear what is said in a group conversation with at least three other people.
	3	Able to hear what is said in a conversation with one other person in a quiet room with a hearing aid, and able to hear what is said in a group conversation with at least three other people, with a hearing aid.
	4	Able to hear what is said in a conversation with one other person in a quiet room, without a hearing aid, but unable to hear what is said in a group conversation with at least three other people even with a hearing aid.
	5	Able to hear what is said in a conversation with one other person in a quiet room with a hearing aid, but unable to hear what is said in a group conversation with at least three other people even with a hearing aid.
	6	Unable to hear at all.
SPEECH	1	Able to be understood completely when speaking with strangers or friends.

- | | | |
|------------|---|---|
| | 2 | Able to be understood partially when speaking with strangers but able to be understood completely when speaking with people who know me well. |
| | 3 | Able to be understood partially when speaking with strangers or people who know me well. |
| | 4 | Unable to be understood when speaking with strangers but able to be understood partially by people who know me well. |
| | 5 | Unable to be understood when speaking to other people (unable to speak at all) |
| AMBULATION | 1 | Able to walk around the neighbourhood without difficulty, and without walking equipment. |
| | 2 | Able to walk around the neighbourhood with difficulty; but does not require walking equipment or the help of another person. |
| | 3 | Able to walk around the neighbourhood with walking equipment, but without the help of another person. |
| | 4 | Able to walk only short distances with walking equipment, and requires a wheelchair to get around the neighbourhood. |
| | 5 | Unable to walk alone, even with walking equipment. Able to walk short distances with the help of another person, and requires a wheelchair to get around the neighbourhood. |
| | 6 | Cannot walk at all. |
| DEXTERITY | 1 | Full use of two hands and ten fingers. |
| | 2 | Limitations in the use of hands or fingers, but does not require special tools or help of another person. |
| | 3 | Limitations in the use of hands or fingers, is independent with use of special tools (does not require the help of another person). |
| | 4 | Limitations in the use of hands or fingers, requires the help of another person for some tasks (not independent even with use of special tools). |
| | 5 | Limitations in use of hands or fingers, requires the help of another person for most tasks (not independent even with use of special tools). |
| | 6 | Limitations in use of hands or fingers, requires the help of another person for all tasks (not independent even with use of special tools). |
| EMOTION | 1 | Happy and interested in life. |
| | 2 | Somewhat happy. |
| | 3 | Somewhat unhappy. |
| | 4 | Very unhappy. |

- 5 So unhappy that life is not worthwhile.
- COGNITION
- 1 Able to remember most things, think clearly and solve day-to-day problems.
- 2 Able to remember most things, but have a little difficulty when trying to think and solve day-to-day problems.
- 3 Somewhat forgetful, but able to think clearly and solve day-to-day problems.
- 4 Somewhat forgetful, and have a little difficulty when trying to think or solve day-to-day problems.
- 5 Very forgetful, and have great difficulty when trying to think or solve day-to-day problems.
- 6 Unable to remember anything at all, and unable to think or solve day-to-day problems.
- PAIN
- 1 Free of pain and discomfort.
- 2 Mild to moderate pain that prevents no activities.
- 3 Moderate pain that prevents a few activities.
- 4 Moderate to severe pain that prevents some activities.
- 5 Severe pain that prevents most activities.

APPENDIX B: SF-6D utility index (Quality Metric)

The SF-6D is a “newer” multi-attribute health state classification and utility scoring system based on 7 dimensions of the SF-36 and provides a means for calculating an indirect community-preference-based “utility” based on individual responses to the SF-36v2. Results from the SF-36v2 are used to estimate a preference-based single index measure for health using a scoring algorithm developed by Brazier and colleagues. (26) The SF-6D was developed by reducing the 8 health domains of the SF-36, to 6 dimensions: physical functioning, role participation (combined role-physical and role-emotional), social functioning, bodily pain, mental health, and vitality. The general health domain was excluded. Individual respondents can be classified on any of four to six levels of functioning or limitations, on each of six domains, thus allowing a respondent to be classified into any of 18,000 possible unique health states to create an “index” score from 0.30 (worst health state) to 1.00 (best health state). A set of preference weights was obtained from a sample of the UK general population (n=836) who were asked to value a sample of 249 possible health states described by the SF6D. Standard Gamble was used as a valuation technique to elicit utility values using a two-stage cascade technique. Upper and lower anchors of dysfunction were assessed by all subjects (all-best SF6D health state (perfect health) and the all-worst SF6D health state). A linear additive regression model was fit by ordinary least squares with SF6D item levels and interactions as covariates. The final model with an R^2 of 0.53, was chosen on the basis of goodness of fit and parsimony. (32)

Short-Form 36 (Psychometric measure-Quality Metric) (SF-36v2):

The SF-36 is one of the most widely used psychometric measures of HRQoL. (12) It is a generic, multi-purpose health status instrument that is intended to measure general health concepts not specific to any age, disease or treatment group. It uses 36 items assesses 8 health dimensions: Physical Functioning, Physical Role, Bodily Pain, General Health, Vitality, Social Functioning, Emotional Role, and Mental Health. The SF-36v2 is a valuable tool in assessing general physical and mental health in populations. For each dimension, item scores are coded, summed, and transformed into a scale from 0 to 100, where 100 is the best possible rating. The SF-36v2 yields results on three levels: (1) item-by-item, (2) an 8-scale profile of functional health and well-being scores that aggregates 2–10 items per domain, and (3) two summary measures (Physical Component Summary Score (PCS) and Mental Component Summary Score (MCS)) that aggregate the 8 scales. Each item is used in scoring only one scale. The aggregated summary measures PCS/MCS use 35 of the 36 items in their summaries (change in general health over past 12 months is not included in summary scores). To obtain summary scores for the physical and mental components, the SF-36 uses a simple arithmetic aggregation across dimensions and assigns them all equal weight in the total score. The dimension scores are calibrated so that the population norm is built into the scoring algorithm, producing a norm-based score (T score transformation with Mean=50 \pm SD=10) based on a 1998 general U.S. population. General validity is accepted due to the widespread use of the SF-36 across a variety of applications. (108)

**Short-Form - 6D (SF-6D) Health State Classification System:
11 of 36 questions from SF-36v2**

Physical Functioning: (3a, 3b, 3j)

The following questions are about activities you might do during a typical day. Does your health now limit you in these activities if so how much?

Vigorous activities: such as running, lifting heavy objects, participating in strenuous sports.

Yes, limited a lot	Yes, limited a little	No, not limited at all
--------------------	-----------------------	------------------------

Moderate activities: such as moving a table, pushing a vacuum cleaner, bowling, or playing golf.

Yes, limited a lot	Yes, limited a little	No, not limited at all
--------------------	-----------------------	------------------------

Bathing or dressing yourself

Yes, limited a lot	Yes, limited a little	No, not limited at all
--------------------	-----------------------	------------------------

Role participation (RP & RE): (4c, 5b)

During the past 4 weeks, how much of the time have you had any of the following with your work or other regular daily activities as a result of your physical health?

Were limited in the kind of work or other activities.

All of the time	Most of the time	Some of the time	A little of the time	None of the time
-----------------	------------------	------------------	----------------------	------------------

During the past 4 weeks, how much of the time have you had any of the following with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

Accomplished less than you would like.

All of the time	Most of the time	Some of the time	A little of the time	None of the time
-----------------	------------------	------------------	----------------------	------------------

Social functioning: (10)

During the past 4 weeks, how much of the time had your physical health or emotional problems interfered with your social activities (like visiting with friends or relatives, etc)?

All of the time	Most of the time	Some of the time	A little of the time	None of the time
-----------------	------------------	------------------	----------------------	------------------

Bodily Pain: (7, 8)

How much bodily pain have you had during the past 4 weeks?

None	Very mild	Mild	Moderate	Severe	Very severe
------	-----------	------	----------	--------	-------------

During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?

None	Very mild	Mild	Moderate	Severe	Very severe
------	-----------	------	----------	--------	-------------

Mental health: (9b, 9f)

These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks...

Have you been very nervous?

All of the time	Most of the time	Some of the time	A little of the time	None of the time
-----------------	------------------	------------------	----------------------	------------------

Have you ever felt downhearted and depressed?

All of the time	Most of the time	Some of the time	A little of the time	None of the time
-----------------	------------------	------------------	----------------------	------------------

Vitality: (9e)

Did you have a lot of energy?

All of the time	Most of the time	Some of the time	A little of the time	None of the time
-----------------	------------------	------------------	----------------------	------------------

SF6D Utility Instrument Health Domains and Domain Levels

Level	Physical Function
1	Not limited vigorous activities
2	Limited a little vigorous activities
3	Limited a little moderate
4	Limited a lot moderate
5	Limited little bathing & dressing
6	Limited a lot bathing & dressing
Level	Role Limitations
1	No role problems
2	Limited in work by physical
3	Accomplish less due to emotional
4	Limited in work and accomplish less
Level	Social Activities
1	Health limits social activities none of the time
2	Health limits social activities a little of the time
3	Health limits social activities some of the time
4	Health limits social activities most of the time
5	Health limits social activities all of the time
Level	Pain
1	No pain
2	Pain but does not interfere work
3	Pain interferes with work a little bit
4	Pain interferes with work moderately
5	Pain interferes with work quite a bit
6	Pain interferes with work extremely
Level	Mental functioning
1	Not tense or downhearted or low any of the time
2	Tense or downhearted or low a little of the time
3	Tense or downhearted or low a some of the time
4	Tense or downhearted or low a most of the time
5	Tense or downhearted or low a all of the time
Level	Vitality
1	Have a lot of energy all the time
2	Have a lot of energy most the time
3	Have a lot of energy some the time
4	Have energy a little of the time
5	Have energy none of the time

APPENDIX C: Quality of Life in Neurological Disorders (Neuro-QoL- Short Form) (10,11,67)

Neuro-QoL is comprised of a set of self-reported item banks and scales that evaluate symptoms, concerns and issues common to a range of neurological conditions. It is designed to generate clinically relevant and psychometrically robust health-related quality of life assessment data in individuals with neurological conditions but does not produce a summary measure of HRQoL; rather, domains are considered independently. The adult form is composed of 17 health related quality of life domains, divided by physical, mental and social overarching domains. The Adult Neuro-QoL measured domains/sub-domains included in LINC cross-sectional population study are: depression; anxiety; fatigue; upper extremity function; lower extremity function; applied cognition-executive function; applied cognition-general concerns; emotional & behavioral dyscontrol; positive affect and well-being; ability to participate in social roles and activities; satisfaction with social roles and activities; stigma; and communication. In the short form version of the Neuro-QoL, each domain has 8-9 items; each item has 5 response levels. A higher Neuro-QoL score in an attribute represents *more of the concept* being measured; not always better status in the attribute. The items on the NeuroQoL were developed with input from various sources; including a literature review, expert interviews, patient and caregiver focus groups, and individual and proxy interviews. Items were further calibrated using Item Response Theory. The tool was then field tested in three waves. Due to the novelty of this tool, the psychometric properties of the NeuroQoL have primarily been investigated in comparison the Quality of Life in Epilepsy (QOLIE-31) scale. Authors indicate that a national normative study is scheduled to be completed in 2011. Initial investigations suggest that the NeuroQoL has demonstrated good convergent validity in relation to the QOLIE-31 (details are described in Nowinski et al., 2010). (67)

Neuro-QoL Instrument Adult Domain Definitions (89)

Domain	Definition
Anxiety	Unpleasant thoughts and/or feelings related to fear. Helplessness, worry and hyperarousal (e.g. tension nervousness and restlessness)
Depression	Experience of loss and feelings of hopelessness, negative mood (e.g., sadness, guilt), decrease in positive affect (e.g., loss of interest), information-processing deficits (e.g., problems with decision-making), negative view of the self(e.g., self-criticism, worthlessness), and negative social cognition
Fatigue	Sensations ranging from tiredness to overwhelming, debilitating and sustained sense of exhaustion that decreases one's capacity for functional, social and mental activities.
Upper Extremity function	One's ability to carry out various activities involving, digital, manual reach-related functions, ranging from fine motor to self-care (activities of daily living).

Domain	Definition
Lower Extremity Function	One's ability to carry out various activities involving the trunk region and increasing degrees of bodily movement, ambulation, balance or endurance.
Applied Cognition-Executive Function	Perceived difficulties in applications of mental function related to planning, organizing, calculating, working with memory and learning
Applied Cognition – General Concerns	Perceived difficulties in everyday cognitive abilities such as memory attention and decision-making.
Emotional & behavioral dyscontrol	A set of disease and/or treatment manifestations including disinhibition, emotional lability, irritability, impatience, and impulsiveness.
Positive Affect and wellbeing	Aspects of a person's life that relates to a sense of well-being, life satisfaction or an overall sense of purpose and meaning.
Sleep Disturbance	Perceptions of sleep quality, sleep depth, restoration of sleep; perceived difficulties with falling asleep, staying asleep.
Ability to participate in social roles and activities	Degree of involvement in one's usual social roles, activities and responsibilities, including work, family, friends and leisure.
Satisfaction with Social Roles and Activities	Satisfaction with involvement in one's usual social roles, activities and responsibilities, including work, family, friends and leisure.
Stigma	Perceptions of self and publically enacted negatively, prejudice, and discrimination as a result of disease-related manifestations

APPENDIX D: Measuring Agreement Between Instruments

The level of agreement between instruments is commonly assessed using methods such as paired t-test of the means, correlations, assessment of variance describing ceiling and floor effects, Intraclass Correlations (ICC), and Bland Altman plots.

1. Means, correlations and variance:

Initially, agreement is assessed comparing means, medians, variance and correlations between measures.

- i. **Means:** Paired t-tests of means are often poor assessments of the level of agreement between instruments since lack of significant differences does not imply agreement between instruments. (28,33,44,50,55)
- ii. **Correlations:** HRQoL measures designed to measure similar constructs will be correlated, but may not necessarily agree sufficiently to be interchangeable. Both means and moderate Pearson or Spearman correlations may mask significant discordance in distributions and agreement overall. An intraclass correlation (ICC) is commonly reported and advocated as a method to assess agreement or consensus between two or more evaluation methods on the same set of subjects. (100) The standard definition of the ICC is given by:

$$\rho = \frac{\sigma_{\alpha}^2}{\sigma_{\alpha}^2 + \sigma_{\epsilon}^2}$$

where σ_{α}^2 represents the between group difference and σ_{ϵ}^2 represents within group differences. The ICC estimates the proportion of *between* subject variation in relation to total variation where 1 represents perfect agreement and 0 - no agreement at all. For instance, if two utility instruments have an ICC value of 0.28, it suggests that only 28% of the variation in utility scores is attributed to *between* subject differences and the remaining 72% is due to variation *within* subjects suggesting poor agreement in the measures. The ICC is criticized for being sensitive to heterogeneity in a sample. A sample with large variance *between* subjects may diminish the effect of within subject differences in ICC calculations inflating suggested agreement between measures in the presence of sample heterogeneity. (54)

- iii. **Variance - Ceiling and floor effects:** Ceiling and floor effects occur when an instrument possesses a distinct upper limit (ceiling) or lower limit (floor) for potential scores, resulting in a concentration of participants scoring at or near this limit. (49) The presence of severe floor and/or ceiling effects indicates the range of data and full health experiences captured are constrained. Ceiling and floor effects can also affect agreement between instruments, particularly at ends of the distributions. Despite the acknowledgement that ceiling and floor effects exist in HRQoL measures – there is no consensus on a definition of what constitutes a ceiling/floor effect, or the measurement of ceiling/floor effects. (49) In cross-sectional data, ceiling and floor effects are often assessed

by quantifying the proportion of respondents with the best possible or worst possible health status. (33,34,49,51,104)

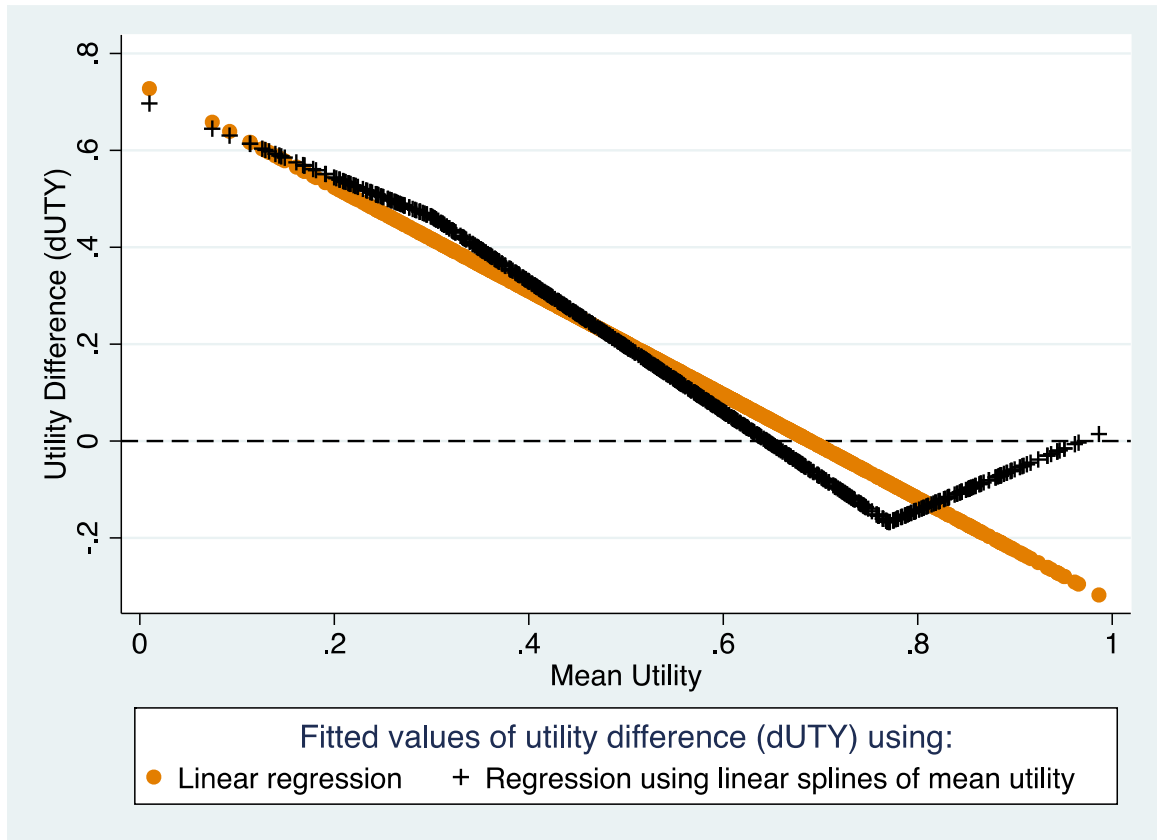
2. **Bland Altman Plots**

Bland Altman Plots (or Tukey mean-difference plots) are proposed as superior to the ICC to identify systematic measurement bias between utility instruments. (59,60) Bland Altman plots are used to visually identify the relationship between the measurement error and best estimate of the true value being measured; indicating if one method consistently over or under estimates results and the association between discordance and the level of utility. (59,60,109) To construct a Bland Altman Plot, the difference between two measures is plotted in relation to the average of the two measures, providing a proportional assessment of the measurement bias. (59) A Bland Altman Plot comparing the HUI3 and SF-6D, for instance, would be constructed by plotting the data points:

$$S(y, x) = (SF6D - HUI3), \frac{SF6D + HUI3}{2}$$

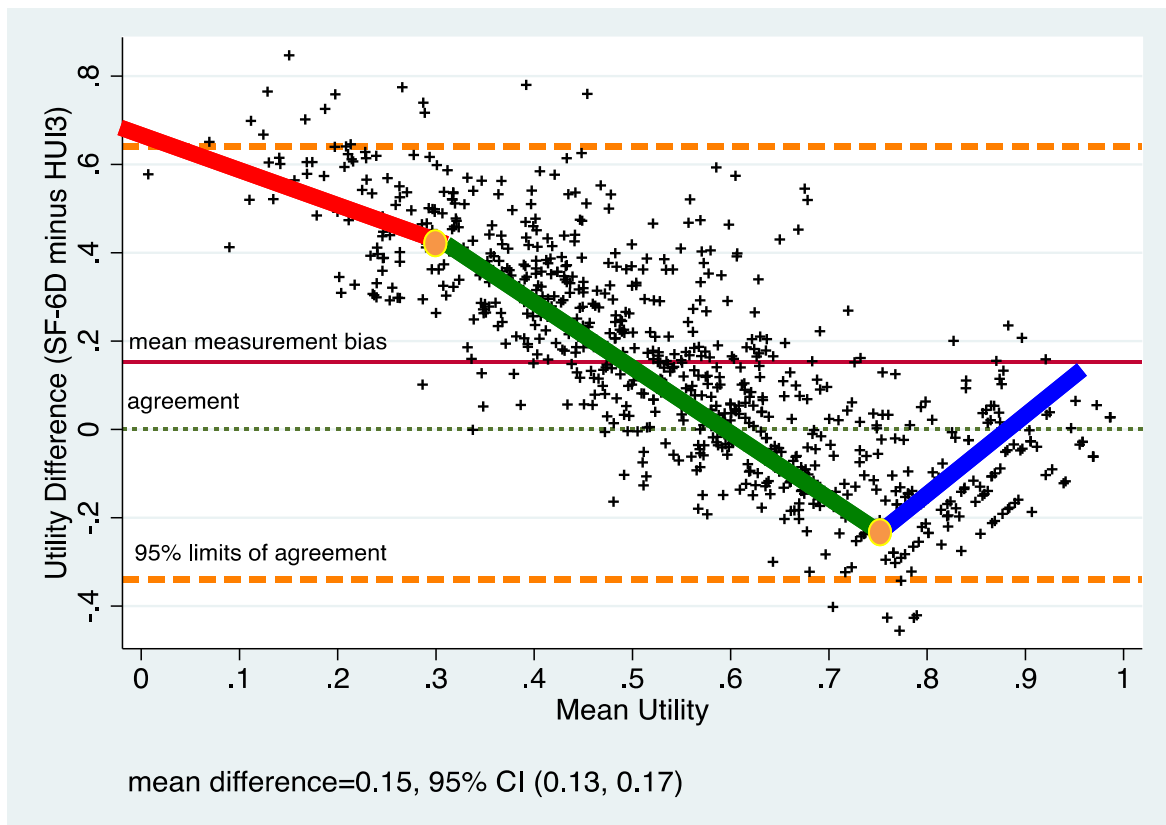
The y-axis is the difference between scores. The x-axis is the mean of the two scores, which serves as a measure of the “true” value being measured. 95% limits of agreement ($\text{mean} \pm 1.96 \text{ SD}$) are added to the plot to provide a visual judgment of how well the measures agree. If the measures generally agree across the range of utility the data points will hover around zero, scattered arbitrarily with no consistent pattern. If discordance is evident the mean difference, referred to as the “bias”, will deviate from zero. The smaller the range between two limits, the better the agreement, though clinical judgment is required to assess whether these limits of agreement (narrow or wide) are clinically significant. If discordance is evident, assessment of the degree of disagreement, and its association with the mean utility score (x-axis), are important. Large standard deviations and widely spaced limits of agreement suggest serious levels of disagreement. (59)

APPENDIX E: Fitted values of dUTY by different regression techniques



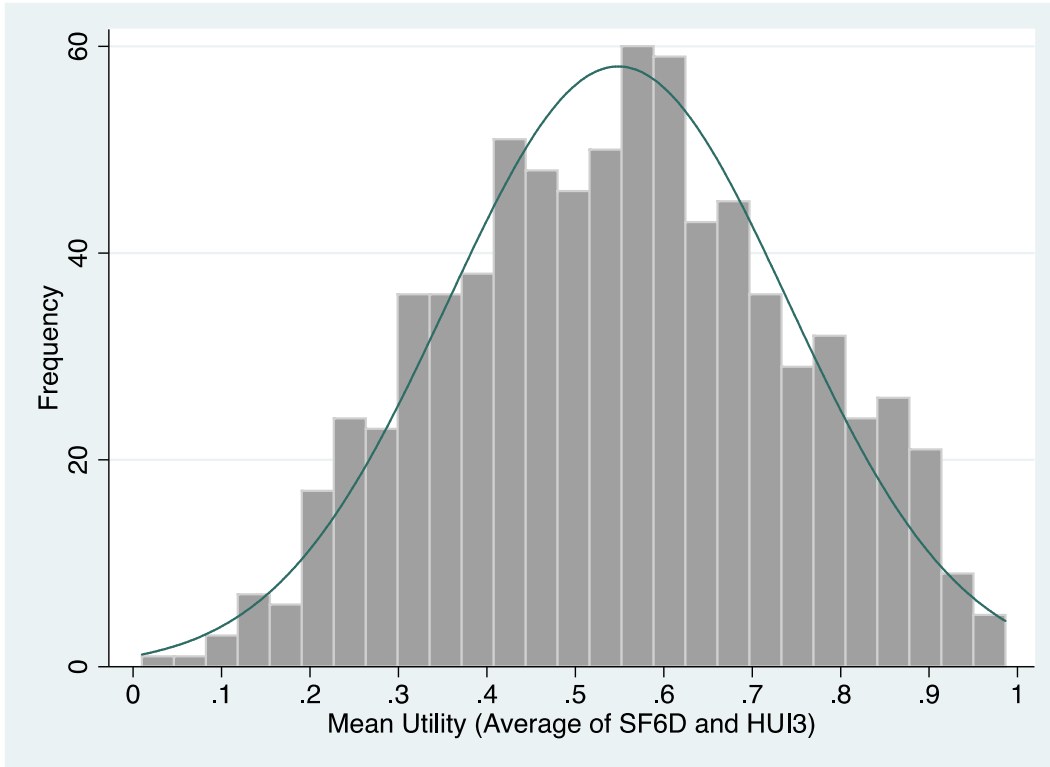
NOTE: The fitted values using simple regression fail to capture the functional form of utility difference in mean utilities between 0.60 and 1.00. Fitting the model using splines captured the change in direction of the regression slope in this utility range.

APPENDIX F: Graphical representation of Regression model using 3 linear splines



Each segment represents a linear spline of mean utility. The orange dots are the “knots” that define the linear segments

APPENDIX G: Distribution of mean utility



Mean utility is calculated as: $\text{mean utility} = \frac{SF6D+HUI3}{2}$