PSYCHOMETRICS OF MEASURES IN AMYOTROPHIC LATERAL SCLEROSIS

THE PSYCHOMETRIC PROPERTIES OF GENERIC-PREFERENCE BASED MEASURES IN AMYOTROPHIC LATERAL SCLEROSIS

By NICOLE PETERS, B.Sc. Honours.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Master of Science

McMaster University

McMaster University MASTER OF SCIENCE (2020) Hamilton, Ontario

(Rehabilitation Science)

TITLE: The Psychometric Properties of Generic Preference-Based Measures in Amyotrophic Lateral Sclerosis AUTHOR: Nicole Peters, B.Sc. Honours (University of Guelph)

SUPERVISOR: Dr. Ayse Kuspinar, PT, PhD. NUMBER OF PAGES: xvii, 142

Lay Abstract

Amyotrophic Lateral Sclerosis (ALS) is a fatal disease that causes individuals to lose their strength and eventually the ability to speak, eat, move and breathe. Questionnaires can be used to understand the health-related quality of life (HRQL) of individuals with ALS however these measures do not always reflect the experiences of these individuals. The goal of this dissertation was to identify whether measures truly capture areas important to individuals with ALS. In our studies, we found that there is little proof in the accuracy of measures used. In addition, the measures do not fully capture the areas of life important to individuals with ALS. This is important to help researchers and health care professionals understand the effects of ALS on HRQL. These results will help them determine which treatments are worthwhile and the best to use in practice and provide recommendations for future research.

Abstract

Background: Amyotrophic Lateral Sclerosis (ALS) is a fatal neurodegenerative disease characterized by the loss of motor neurons. Preference-based measures (PBMs) of health-related quality of life (HRQL) can be utilized for cost-effectiveness analyses of interventions in individuals with ALS. However, current measures are generic (GPBMs) and the psychometric properties of these measures have not yet been evaluated in ALS.

Purpose: The purpose of this thesis was to evaluate the psychometric properties of GPBMs in ALS by 1) conducting a systematic review of the psychometric properties of GPBMs, and 2) assessing the content and convergent validity of GPBMs in ALS.

Methods: Two studies were conducted. First, a systematic review was performed, and four databases were searched to identify studies that used and reported on the psychometric properties of GPBMs in ALS. Second, participants were recruited from three clinical sites across Canada and outcome measures were administered through an online or hardcopy survey. Areas of importance to the HRQL of individuals with ALS were identified using the Patient Generated Index (PGI), mapped against GPBMs to determine their coverage and scores were compared to determine convergent validity.

Results: For the first study, the EQ-5D-3L was found to be the most commonly used GPBMs in ALS. It demonstrated convergent and known-groups validity however, significant floor effects were observed. For the second study, results indicated that the majority of GPBMs identified approximately half of the areas impacted by ALS. In addition, there were several domains not identified by GPBMs.

Conclusion: This thesis highlights the importance of complete psychometric evaluation of measures in ALS. There is the need for the development of an ALS specific preference-based

measures that reflects the health concerns of individuals with ALS; as GPBMs used in ALS were evaluated and deemed to be lacking in support for their usage in ALS.

Acknowledgments

This would not have been possible without the tremendous support I have received throughout the duration of my Master's degree at McMaster University. These last two years have enabled my potential as a researcher to grow and develop and has provided me with extensive opportunities and challenges.

I would like to express my eternal gratitude to my supervisor, Dr. Ayse Kuspinar, for her unwavering support throughout my graduate studies. I have been fortunate to have her as a supervisor and mentor; and could not have wished for a more kind and supportive person to work with. The continuous backing, guidance and encouragement I have received from her has improved my learning and molded my critical thinking. Ayse, I cannot express how your kindness, understanding and openness has helped these past two years. Thank you.

An immense thank you to my committee members, Dr. Vanina Dal Bello-Haas and Dr. Tara Packham. They have provided critical feedback and insights that have enhanced the overall quality of my work. Thank you for always being available to answer questions and providing me with honest and critical feedback that has helped enhance my academic writing skills.

Thank you to my fellow colleagues and friends, and especially everyone who has aided my research. I have worked, officially and unofficially, with many of you these past couple of years and the input from everyone was invaluable. A special thanks to my research team and fellow colleagues who have contributed to my thesis and have helped me prepare for my final defense. Lastly, I would like to thank my parents for pushing me to continue my studies and support me all these years. Special thanks go out to my friends and family. I am truly fortunate to have so many loving people in my life, who support me and provide me with encouragement and positivity.

Table of Contents

List of Figures and Tables

Chapter Two

Chapter Three

List of Abbreviations

15D	15 Dimension
95% CI	95% Confidence interval
ALS	Amyotrophic Lateral Sclerosis
ALS/HSS	Amyotrophic Lateral Sclerosis Health State Scale
ALSAQ-40	Amyotrophic Assessment Questionnaire-40
ALSFRS-R	Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised
ALSSQOL-R	Amyotrophic Lateral Sclerosis Specific Quality of Life-Revised
ALSSS	Amyotrophic Lateral Sclerosis Severity Scale
AQoL-8D	Assessment of Quality of Life-8 Dimension; other versions include the 4D,6D,7D
AUC	Area under the curve
BDI	Beck Depression Inventory
CALS	Canadian ALS Research Network
CINAHL	Cumulative Index to Nursing and Allied Health Literature
COSMIN	Consensus-based Standards for the section of health Measurement INstruments
EQ-5D	EuroQol 5 Dimension
EQ-5D-3L	EuroQol 5 Dimension 3 Level
EQ-5D-5L	EuroQol 5 Dimension 5 Level
EUT	Expected utility theory
FALS	Familial Amyotrophic Lateral Sclerosis
FSHD	Facioscapulohumeral Muscular Dystrophy
FTD	Frontotemporal dementia
FVC	Forced vital capacity

GPBM	Generic preference-based measure
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HRQL	Health-related quality of life
HUI/1	Health Utilities Index/Mark 1
HUI2	Health Utilities Index Mark 2
HUI3	Health Utilities Index Mark 3
ICF	International Classification of Functioning, Disability and Health
LMN	Lower motor neurons
MAU	Multi-attribute utility
MCS	Mental Component Summary score of the SF-36
MCS-12	Mental Component Summary score of the SF-12
MeSH	Medical subject heading
MG	Myasthenia Gravis
MiToS	Milano-Torino staging system
MND	Motor neuron disease
MS	Multiple Sclerosis
n	Frequency of appearance
p	Probability
PBM	Preference-based measure
PCS	Physical Component Summary score of the SF-36
PCS-12	Physical Component Summary score of the SF-12
PGI	Patient Generated Index
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PRO	Patient-reported outcome
QALY	Quality-adjusted life years
QOL	Quality of life
QWB	Quality of Well-Being scale
QWB-SA	Quality of Well-Being Self-Administered scale
r	Pearson's correlation
SD	Standard deviation
SEIQoL	Schedule for the Evaluation of Individual Quality of Life
SEIQoL-DW	Schedule for the Evaluation of Individual Quality of Life - Direct Weighting
SF-12	Short Form-12
SF-36	Short Form-36
SF-6D	Short Form 6 Dimension; SF-36 or SF-12 derived
SG	Standard gamble
SIP	Sickness Impact Profile
SIP/ALS-19	Sickness Impact Profile Amyotrophic Lateral Sclerosis-19
TTO	Time trade off
UMN	Upper motor neurons
VAS	Visual analogue scale
WHO	World Health Organization

Declaration of Academic Achievement

The following is a declaration of interest that the content of the research in this document has been completed by Nicole Peters. It recognizes the contributions of Dr. Vanina Dal Bell-Haas, Dr. Tara Packham and Dr. Ayse Kuspinar in components of the research process and completion of this thesis. Nicole Peters is the primary author and contributor of this thesis. Preliminary edits were made by Dr. Ayse Kuspinar and all authors commented on previous iterations of the manuscripts. All authors contributed to the studies' conceptions and designs. Development of the surveys, data collection and analyses were completed by Nicole Peters. Ava Mehdipour and Jill Van Damme contributed to components of data collection and analyses in each of the studies, respectively. Dr. Vanina Dal Bello-Haas and Dr. Tara Packham provided feedback and contributed to all chapters of this thesis.

Chapter One: Introduction

1.1 Amyotrophic Lateral Sclerosis

Amyotrophic lateral sclerosis (ALS), also known as Lou Gehrig's disease, is a fatal neurodegenerative disease characterised by selective and progressive degeneration of motor neurons of the brain (upper motor neurons; UMNs) and spinal cord (lower motor neurons; LMNs) (Zarei et al., 2015). In addition to UMNs and LMNs, neurons in the frontal cortex and other neuroanatomical regions may also be affected. ALS was first described in 1869 by Jean-Martin Charcot (Charcot & Joffroy, 1869) and is considered the most common motor neuron disease (MND). Most cases of ALS (90-95%) are sporadic ALS, meaning the cause of the disease is unknown (although genetics could be involved) (Pasinelli & Brown, 2006). The remaining 5-10% of cases are considered familial ALS (FALS), meaning the cause of the disease is likely due to a dominant genetic mutation (Pasinelli & Brown, 2006).

ALS affects more than 200,000 people worldwide, has an incidence rate of 0.5 per 100,000 person-years in Canada (Longinetti & Fang, 2019); and, most commonly develops in people between the ages of 30 to 80 years, with a mean age of onset ranging from 55 to 66 years (Longinetti & Fang, 2019). Additional suspected risk factors include male sex, smoking, exposure to occupational toxicants and physical activity (Ingre et al., 2015; Longinetti & Fang, 2019). Several positive prognostic factors have been reported such as spinal onset (Leighton et al., 2019), younger age at onset and higher baseline functional scores (Karanevich et al., 2018). Individuals with ALS experience wide variability in disease severity and disease progression with an overall mortality rate of 80% within the first 2-5 years of diagnosis (Longinetti & Fang, 2019; Valko & Ciesla, 2019). Currently, there is no cure for ALS and no effective treatment to halt or reverse its progression, although there are two disease-modifying agents, riluzole and

edaravone, that have been approved for treatment of ALS (Canadian Agency for Drugs and Technologies in Health, 2019; Schultz, 2018).

1.1.1 Signs and Symptoms

There are three forms of ALS (Kiernan et al., 2011): 1) limb-onset ALS, 2) bulbar-onset ALS, and 3) respiratory-onset ALS. Limb-onset ALS is the most common type of ALS, accounting for 70% of cases. Bulbar-onset ALS is observed in 25% of cases, with the remaining 5% of cases having initial trunk or respiratory involvement without significant limb or bulbar symptoms (Hardiman et al., 2011; Kiernan et al., 2011).

The clinical hallmark of ALS is the presence of both UMN and LMN involvement (Kiernan et al., 2011). Loss of LMNs results in muscle weakness, atrophy, cramps and fasciculations (muscle twitching). Loss of UMNs contribute to spasticity (muscle stiffness), and hyperreflexia (exaggerated reflexes).

The onset of ALS can be subtle, with gradual development of signs and symptoms. For many individuals with limb-onset ALS, localized muscle weakness and atrophy in the upper or lower limbs are the first symptoms to appear (Wijesekera & Leigh, 2009). Individuals will experience initial difficulties performing simple tasks such as buttoning a shirt, if upper limb onset, or awkwardness when walking or running, if lower limb onset. Regardless of where symptoms first appear, muscle weakness and atrophy will begin to affect other areas of the body as the disease progresses and degeneration of motor neurons spread. Eventually, individuals will develop bulbar and respiratory symptoms, such as problems with swallowing (dysphagia), speaking or forming words (dysarthria) and breathing (dyspnea) (Wijesekera & Leigh, 2009).

Patients with bulbar-onset ALS present with dysphagia or dysarthria resulting from weakness in the muscles of the face and throat (Wijesekera & Leigh, 2009). Similar to limb-

onset ALS, as degeneration spreads, muscle weakness and atrophy will affect other areas of the body such as the limbs. Sialorrhea (excessive drooling), resulting from difficulty swallowing, and pseudobulbar symptoms such as emotional liability and excessive yawning also commonly develop (Wijesekera & Leigh, 2009). The remaining 5% of cases present with respiratory weakness without significant limb or bulbar symptoms. The patients in this group display symptoms such as dyspnea, shortness of breath (orthopnea) and disturbed sleep; or nocturnal hypoventilation (shallow breathing) (Zarei et al., 2015).

Irrespective of the site of disease onset, as the muscles of the respiratory system weaken, individuals will lose the ability to breath on their own and require ventilation to survive. Thus, during late stages of the disease, a combination of degeneration symptoms appear that will ultimately result in death; most often due to respiratory failure or pulmonary complications (Bäumer et al., 2014; Gordon, 2013; Silani et al., 2011).

The majority of individuals diagnosed with ALS retain their sensation, eye movement, bowel, bladder and sexual functions (Kiebert et al., 2001; Smith et al., 2000). Cognitive impairment is present in more than 40% of patients with ALS, and approximately 5 to 15% of patients develop frontotemporal dementia (FTD) (Phukan et al., 2012). Furthermore, ALS has an impact on the psychological, emotional, and mental health of individuals (Lou et al., 2003). Throughout the progression of ALS, research has ascertained the effects of ALS on the psychological and mental health of individuals. Therefore, health-related quality of life (HRQL) is strongly influenced by not only physical but also mental health of individuals with ALS (Prell et al., 2019; Robbins et al., 2001; Simmons, 2015; Simmons et al., 2000; van Groenestijn et al., 2016; Zarei et al., 2015).

1.1.2 Diagnosis

As the cause of ALS is unknown, there is no known biological marker. Therefore, the diagnosis of ALS is primarily a clinical diagnosis that requires: 1) clinical confirmation of UMN and LMN involvement, 2) presence of disease progression, and 3) exclusion of potentially similar conditions that can mimic ALS (Hardiman et al., 2011; Turner & Talbot, 2013). Standard clinical criteria and diagnostic tests, including patient history, physical examination, laboratory studies, electrodiagnostic tests, neuroimaging and potential genetic testing are used in the diagnosis of ALS (Hardiman et al., 2011).

The El Escorial criteria for diagnosing ALS was developed in 1994 by the World Federation of Neurology for research and trial purposes (Brooks, 1994). It was revised and renamed the Airlie House criteria in 1998 in order to incorporate laboratory testing (Brooks et al., 2000; van den Berg et al., 2019) and recognize the role of neurophysiology in diagnostic criteria (Schrooten et al., 2011). The El Escorial criteria and Airlie House criteria are used to both predict the degree of certainty of ALS diagnosis and as inclusion criteria for research study/trial classification of ALS patients (Ludolph et al., 2015). The El Escorial and Airlie House criteria's consists of four categories of ALS: definite ALS, probable ALS, possible ALS and suspected ALS (Brooks, 1994; Ludolph et al., 2015). A definite diagnosis of ALS requires evidence of LMN and UMN degeneration and progression of muscle weakness, within a region or to other regions. Additionally, there must be the absence of electrophysical, pathological and neuroimaging evidence of other disease processes that might explain the symptoms and signs (Brooks, 1994). However, the criteria was not designed for everyday clinical practice and has been criticized as being excessively restrictive (Traynor et al., 2000). Moreover, it does not take into account the nonmotor components of ALS therefore, further revisions are required to modify the system accordingly (Hardiman et al., 2011; Ludolph et al., 2015).

1.1.3 Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised (ALSFRS-R)

The Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised (ALSFRS-R) is a measure most commonly used in clinical practice and clinical research trials. The ALSFRS-R (Cedarbaum et al., 1999) was developed in 1999 in response to critique about the original ALSFRS (Brooks et al., 1996; Cedarbaum & Stambler, 1997). The revised version is an interviewer-administered measure that quantifies degree of functioning in ALS. There is a total of 12 questions across 3 domains: bulbar (3 questions), motor (6 questions) and respiratory (3 questions). The questions are rated on a five-point scale from 0 (complete dependence) to 4 (normal function). A total ALSFRS-R score, from 0 to 48, is produced through summation of the individual item scores; with higher scores indicating full functioning and increased predicted survival (Cedarbaum et al., 1999). The ALSFRS-R is more sensitive to change than the original measure and demonstrates strong internal consistency (Cronbach's $\alpha = 0.73$) and construct validity (Pearson's correlation coefficient = -0.71 with the physical dimension scale of another measure of quality of life (QOL), the Sickness Impact Profile (SIP)) (Cedarbaum et al., 1999). It also demonstrates a better ability to predict survival than the original ALSFRS (Cedarbaum et al., 1999).

In addition to the interviewer-administered measure, a self-administered version of the ALSFRS-R was developed to reduce patient burden resulting from frequent visits to the clinic and to allow patients to monitor disease progression from home (Montes et al., 2006). Similar to the interview administered version, the self-administered ALSFRS-R consists of 12 questions across 3 domains: bulbar (3 questions), motor (6 questions) and respiratory (3 questions). It produces both a total score, from 0 to 48, and 3 subdomain scores; with higher scores indicating full functioning. The self-administered ALSFRS-R has demonstrated excellent reliability (intra-

class correlation=0.93, 95% CI 0.88 to 0.96) and sensitivity to change over time (Montes et al., 2006).

In clinical practice, the ALSFRS-R is used by health care professionals as a measure of disease severity and as a prognostic indicator. According to Kaufmann et al. (2005) and Cedarbaum et al. (1999), change in ALSFRS-R scores over time can be used to measure disease severity and predict survival time; whereby one's mean survival time is predicted to be greater when the total score is ≥ 38 (out of 48).

Studies have evaluated the psychometric properties of the ALSFRS-R using Rasch analysis (Franchignoni et al., 2013; Franchignoni et al., 2015) and longitudinal and survival analyses (Rooney et al., 2017). Due to the lack of unidimensionality (when a single construct underlying the items is measured) (Bond & Fox, 2015) and the presence of differential item functioning (when items have significantly different meanings for different groups, despite equal levels of the characteristic being measured) (Bond & Fox, 2015; Pallant & Tennant, 2007) between patients with limb versus bulbar onset, researchers caution the reporting of a single total score; and recommend reporting domain specific subscale scores organized into bulbar, motor and respiratory domains (Cedarbaum et al., 1999; Franchignoni et al., 2013; Franchignioni et al., 2015; Rooney et al., 2017). A total subscale score for each domain is produced through summation of the corresponding items: items 1 to 3 (bulbar), items 4 to 9 (motor) and items 10 to 12 (respiratory) (Franchignoni et al., 2013). Bulbar and respiratory domain scores range from 0 to 12, while the motor domain scores range from 0 to 24; with higher scores indicating full functioning.

1.1.4 Treatment and Care

Currently, riluzole and edaravone are the only two evidence-based disease modifying drugs approved for ALS (Schultz, 2018; Zarei et al., 2015). Management of ALS is focused on symptom control, enhancing function and preservation of QOL (Hardiman et al., 2011). As the most common cause of death in ALS is respiratory failure, focus is given to respiratory assessment and management using non-invasive ventilation. Additionally, attention to palliative care supports and end-of-life decisions, such as by providing patients with a realistic projection of disease trajectory and consideration of an advanced directive, are required due to the rapid nature of ALS (Bede et al., 2011).

1.2 Health and Quality of Life (QOL)

The terms 'health' or 'health status', 'QOL' and 'HRQL' are used in the literature to define different things by different instrument developers. Given that these are key terms, and that QOL is impacted by ALS, clear and appropriate use of each is important; to both provide a key distinction of which construct an instrument purports to measure and to avoid potential for confusion.

A long-standing definition from the World Health Organization (WHO) (1958) defines health as "a state of complete physical, mental and social well-being, and not merely the absence of disease and infirmity." Of importance in this definition is the inclusion of social well-being and emphasis on health as more than the absence of disease (Karimi & Brazier, 2016). While this definition has been used in the development of various measures, it is very broad and not easy to operationalize. Furthermore, in clinical and health services research, health has been defined more so only as an absence of illness, disease and injury (Peasgood et al., 2014). Health status is

thus more than performance of activities of daily living; it considers broader medical and functional well-being and is sometimes reported in terms of impact of disability (Bergner, 1989).

With assessments of QOL and HRQL there is a shift in the emphasis on well-being to a subjective measure, with the inclusion of outcomes not directly observable or characterized by others (Fayers & Machin, 2016; Rosenbaum et al., 2007). Although the domains captured in QOL or HRQL measures may be considered subjective or objective, such as the ability to walk or the severity of bodily pain, the ratings of these dimensions are ultimately subjective since they are determined by the individual (Calman, 1984; Haas, 1999). Contrariwise, some authors have argued that objective factors should be included in measures of QOL and that it is best to include both (Cummins, 2005; Felce & Perry, 1995; Meeberg, 1993). For example, Felce & Perry (1995) define QOL as an overall general well-being that includes both objective and subjective evaluations with the extent of one's personal development and level of purposeful activity, all weighted by a personal set of values.

In general, QOL is a global construct defined as "an individuals' perception of their position in life in the context of their culture in which they live and in relation to their goals, expectations, standards and concerns" (Kuyken & Group, 1995). Experts accept that QOL has been typically characterized as an "umbrella term" that captures aspects of one's life that are outside of the health purview, in addition to life satisfaction and well-being (Peasgood et al., 2014). Calman (1984) defines QOL as "the extent to which our hopes and ambitions are matched by experience" and suggests that the key aim of care should be to "narrow the gap between a patient's hopes and expectations and what actually happens." Both of these definitions emphasize the multidimensional aspect of the physical, psychological, social and spiritual dimensions of QOL (Haas, 1999). Furthermore, it is acknowledged that QOL is influenced by

factors such as material and economic circumstances (i.e. housing, employment, standard of living), freedom and satisfaction with life (Guyatt, 1993).

The concept of QOL is important in both research and clinical practice to evaluate treatment and clinical care, especially among patients with a progressive disease. However, it has been considered too general to be of use in health care since health is only one factor of QOL. Therefore, HRQL is deemed the appropriate focus in health care even though almost all aspects of life can become health-related when disease and illness are experienced by a patient (Guyatt, 1993).

1.3 Health-Related Quality of Life (HRQL)

HRQL is a multidimensional construct that focuses on aspects of one's life within the purview of the health care system. Kaplan (1985) defines HRQL in terms of the impact of disease and treatment on disability and daily functioning. Patrick and Erickson (1993) define HRQL as the value assigned to the duration of life as modified by impairments, functional states, perceptions and social opportunities and as influenced by disease, injury, treatment and policy. Overall, both definitions include issues that are of relevance and importance to an individual's well-being. Therefore, HRQL can be thought of as a multidimensional, patient-reported construct in the health-care purview; i.e. "an individual's perception of how an illness and its treatment affect the physical, mental and social aspects of one's life" (Mayo, 2015). This definition in particular is of importance as HRQL is a subjective measure, whereby descriptions of the experience of a health state would be best elicited from individuals themselves in order to reflect the actual experience of the disease and its treatment (Fitzpatrick et al., 1998).

Therefore, HRQL is a patient-perceived construct, otherwise known as a patient-reported outcome (PRO) (Simmons, 2015). PROs are defined as "any reports coming directly from patients about how they function or feel in relation to a health condition and its therapy, without interpretation of the patients responses by a clinician or anyone else" (de Vet, 2011; Patrick et al., 2007). PROs have increasingly become significant in the health-care field as healthcare providers must consider the impact of the disease on the patient's QOL in order to optimize clinical care (Brazier, Ratcliffe, et al., 2017; Simmons, 2015). This is especially important with chronic conditions and progressive diseases, such as ALS in assessing the impact of the disease and treatment on patients, due to the variability and complexity of the conditions (Guyatt, 1993).

Measures of HRQL can be used to provide a profile of the current health status of an individual over time which allows for treatment decision making and outcome evaluation purposes (Guyatt, 1993). Moreover, improvement in HRQL is often considered the ultimate goal in health-care (Ruta & Garratt, 2013). As HRQL has been shown to be impacted in chronic and progressive illnesses (Ojelabi et al., 2017) as a result of one's physical and psychological well-being (Simmons, 2015; Simmons et al., 2000; van Groenestijn et al., 2016; Zarei et al., 2015), HRQL should be well-understood and conceptualized. As such, researchers have indicated that studies on HRQL should be based on conceptual models that enhance the understanding of linkages and facilitate the design of protocols for optimal care (Bakas et al., 2012). There are many HRQL models, developed for different contexts and purposes, and with different domains and definitions included. However, the three most commonly used HRQL models are (Bakas et al., 2012): 1) the Wilson & Cleary Model of HRQL (Wilson, 1995), 2) the Revised Wilson & Cleary Model of HRQL (Ferrans et al., 2005), and 3) the WHO's International Classification of Functioning Disability and Health (ICF) (World Health Organization, 2001).

The Wilson & Cleary Model of HRQL (Wilson, 1995) is one of the most widely used conceptual frameworks of HRQL. It was the first model developed for health care to integrate clinical (biomedical) approaches with psychosocial approaches and provide a conceptualization of HRQL (Bakas et al., 2012). The model includes five domains (biological and physiological factors, symptoms status, functioning, general health perceptions and overall QOL) that interact with each other along a causal pathway starting with the bio-physiological and moving outwards to the interaction of the individual as a social being (Wilson, 1995). In addition to the five domains, the authors suggest that environmental and individual factors are also associated with the model as outcomes of health, rather than as domains, and thus still affect HRQL (Bakas et al., 2012; Wilson, 1995). The Wilson & Cleary model was developed with clinicians and researchers in mind in order to provide a broader view of HRQL; and to help target, rather than just monitor, the improvement of HRQL in clinical trials (Bakas et al., 2012).

The Revised Wilson & Cleary Model of HRQL was developed by Ferrans et al. (2005) as an expansion of the original model. With the same five domains, the revision explicitly defines individual and environmental factors and incorporates nonmedical factors, such as demographics or interpersonal relationships, into individual or environment factors. In addition, arrows depicting a causal relationship in the figure are removed in order to provide an explicit understanding of the relationships. The revised model could also be applied to any health care discipline, in contrast to the original model which was developed for physicians only (Ferrans et al., 2005).

The International Classification of Functioning, Disability and Health (ICF) (World Health Organization, 2001), developed by the World Health Organization, is a model designed to integrate the biomedical and social approaches of health-care and provide a standard of language

across cultures and disciplines. The ICF is a classification of health and health-related states, conceptualized in terms of functioning and disability and organized according to a set of principles (World Health Organization, 2001). The ICF consists of four key components (labeled as ICF chapters) organized into two parts. Each component can be expressed in both positive and negative terms. Part 1 deals with Functioning and Disability ((b) body functions and (s) structures, (d) activities and participation). Part 2 covers Conceptual Factors ((e) environment factors and personal factors). However, unlike the Wilson & Cleary Model(s) (Ferrans et al., 2005; Wilson, 1995), the ICF is not specific to HRQL (Cieza & Stucki, 2008). Therefore, the ICF can be used as a mapping and classification framework for coding and description of health-related problems across levels of impairments, activity limitations and participation restrictions; rather than as a guide for hypothesis generation in the area of HRQL (Bakas et al., 2012; Kuspinar et al., 2019; Kuspinar & Mayo, 2013; Mayo et al., 2011).

1.4 Outcome Measures

Patient-perceived constructs, i.e. PROs such as QOL or HRQL, have increasingly been used in the literature to assess the efficacy and effectiveness of health care interventions (Brazier, Ratcliffe, et al., 2017). Instruments have been developed to measure these constructs for this reason. In addition to their use in clinical decision-making and evaluation purposes, HRQL measures have been used in conducting economic evaluations for reimbursement decisions (Guyatt, 1993).

1.4.1 Structure of Measures

HRQL measures can differ in their content from generic concepts of functioning for a variety of health conditions through to specific dimensions, such as symptoms, that are relevant

to a specific disease (or condition or population). Thus, these measures can be generic (disease-agnostic) or disease-specific. Disease-specific measures are intended to have very relevant content for the target disease as they have been developed explicitly to assess health problems (Fitzpatrick et al., 1998). They contain domains that are more likely to detect important changes that occur over time in a disease being studied. However, the greatest disadvantage of disease-specific measures is that most measures cannot be used for comparative purposes against a general sample or other condition. Additionally, comparisons cannot be made between outcomes of different treatments, such as for resource allocation, across different health conditions (Fitzpatrick et al., 1998). Generic measures, on the other hand, are designed to assess a broad range of aspects of HRQL with domains relevant for a variety of conditions (Fitzpatrick et al., 1998). This is an advantage as they measure change in health for a variety of conditions, which enables comparisons across treatments and effectiveness purposes. Additionally, generic measures tend to reduce patient burden (e.g. by asking generic questions) and generate normative values in which patients with health conditions can be compared. A substantial disadvantage however is some loss in relevance of the questionnaire when applied to specific conditions, and less sensitivity to changes in a specific health condition if items are not included that are of relevance to that population (Fitzpatrick et al., 1998).

HRQL can be measured using: i) individualized measures, ii) health profiles or iii) preference-based measures (also known as utility measures).

1.4.2 Individualized Measures of HRQL

Individualized measures of HRQL utilise the individual's perspective of HRQL based on areas of life that the individual considers to be of personal concern for him or her at the present time (Joyce et al., 1999). The individual is asked to identify and rate those aspects of life affected

by health without an imposed list of potential answers (Ruta et al., 1994). As perceptions vary between individuals, what is important to one may have little or no relevance to another. Furthermore, it has been suggested that by excluding items not directly of concern to the individual, individualized measures are able to eliminate extraneous 'noise' present in other standardized measures, leading to improved responsiveness to change (Martin et al., 2007). Two such instruments have received attention for eliciting respondent's own concerns and perceptions; the Schedule for the Evaluation of Individual Quality of Life (SEIQoL) – Direct Weighting (SEIQoL-DW) and the Patient Generated Index (PGI).

The SEIQoL (O'Boyle, 1994) was intended to provide an individualized assessment of overall QOL and is administered in the form of a semi-structured interview. There are three parts to the measure. The first part requires the respondents to nominate the five most important areas of life (referred to as elicited cues) related to their QOL. Second, the five areas in addition to their overall QOL, are then rated on a visual analogue scale (VAS) from 'as good as it could be' to 'as bad as it could be'. The last stage requires respondents to respond to 30 hypothetical vignettes which vary in relation to the properties identified as important to them. Weights are produced using multiple regression analysis for the relative importance of each cue to the overall VAS judgement of QOL (O'Boyle, 1994). A single index score from 0 to 100 is produced for the SEIQoL by multiplying the rating and the weights for each of the five elicited cues and summing these products (O'Boyle, 1994). A shorter method, the SEIQoL-Direct Weighting (SEIQoL-DW), was developed in order to introduce a simpler weighting procedure (Hickey et al., 1996). The first two stages of the measure are the same as the original. The third stage involves quantifying the relative contribution of each cue using a scale from 0 to 100 and a corresponding visual pie chart with interlocking and colored disks that represent the five areas nominated. The

respondent is asked to adjust the disks until the size of each colored segment corresponds to the relative importance. The total value of all five weights results in a single score from 0 to 100. Both measures have been found to be valid and reliable and have been used in various conditions such as HIV/AIDS (Hickey et al., 1996), hip-replacement surgery (O'Boyle et al., 1992), cancer and neurological disorders (Wettergren et al., 2009), including ALS (Neudert et al., 2001).

The Patient Generated Index (PGI) (Ruta et al., 1994) is an interviewer or self-administered individualized measure that was developed to focus on the impact of a specific health condition on QOL. Participants are first asked to identify up to five of the most important areas of their HRQL affected by the specified health condition. Next, they are asked to rate the extent of impact of each area on their lives, in addition to 'all other areas of their lives not mentioned', from 0 (the worst you could imagine) to 10 (exactly as you would like it to be). Last, participants are asked to imagine that they could improve some or all their chosen areas. Participants are given twelve spending 'points' and are asked to distribute the points across the areas nominated to indicate the relative importance of each area. More points allocated to an area indicates greater importance and hope of improvement. The rating score and weighting points are then multiplied and summed to produce a single index score of overall HRQL, reported as a percentage, with higher scores indicating better HRQL. The PGI demonstrates validity and responsiveness to change (Martin et al., 2007), and has been used in various neurological conditions such as Parkinson's disease (Kuspinar et al., 2019) and Multiple Sclerosis (MS) (Kuspinar & Mayo, 2013).

1.4.3 Health Profiles

Health profiles, such as the Short Form-36 (SF-36) (Ware et al., 1993) and ALS Specific QOL-Revised scale (ALSSQOL-R) (Felgoise et al., 2011; Simmons et al., 2006), are generic or

specific HRQL instruments that attempt to measure all aspects of HRQL (Guyatt, 1993). They have a simple summative scoring system whereby responses to the dimensions are numerically coded and summed for each dimension. By using a simple summative scoring system, health profiles assume both an equal interval between response levels of an item and an equal importance between the items in a dimension (Brazier, Ratcliffe, et al., 2017). For example, this scoring method does not necessarily reflect the value an individual would place on different items when walking vs. climbing stairs if they lived in a one-story house (Brazier, Ratcliffe, et al., 2017). Additionally, numerous studies have shown that intervals between response choices are not equal and items do not have the same weight; threatening the validity of such health profiles (Brazier, Ratcliffe, et al., 2017).

The SF-36 (Stewart & Ware, 1992; Ware et al., 1993) is one of the most widely used generic health profile measures of HRQL. The SF-36 consists of 35 items across eight dimensions: physical functioning (ten items), role limitation (physical) (four items), bodily pain (two items), general health (five items), vitality (four items), social functioning (two items), role limitation (psychological) (three items) and mental health (five items) (Ware et al., 1993). Each dimension varies in the number of items and number of response choices. Total scores are produced for each dimension in addition to two summary scores, the Physical Component Summary (PCS) score and the Mental Component Summary (MCS) score. Higher scores indicate better health, and scale scores, once transformed, range from 0 to 100. The Short Form-12 (SF-12) (Ware et al., 1996), another health profile of HRQL, is derived from the SF-36, and contains a subset of 12 items covering the same eight domains of health outcomes. It was created to reduce the burden of response for patients as it only takes a few minutes to administer. The measure yields the same two summary scores (PCS-12 and MCS-12) however unlike the SF-36,

it does not provide separate scores for each domain. Both the SF-36 and SF-12 have been shown to reflect change over time (Brazier et al., 1999; Harper et al., 1997; Jenkinson et al., 1997; Walters et al., 1999). In ALS, the SF-36 has shown to be responsive to change, as evidenced by a decline in HRQL and physical function with disease progression (Neudert et al., 2004; Simmons, 2015). However, there is a demonstrated floor effect for both the SF-12 and SF-36 in various conditions, including ALS (Bindman et al., 1990; Jenkinson et al., 2002).

The SIP (Bergner et al., 1981) is another well-known health profile of HRQL. It is a measure of perceived health status and is intended to provide a measure of the effects or outcomes of health care. The SIP consists of 136 items divided into 12 subscales and takes about 20-30 minutes to complete. However, there is no global question about overall health or QOL. The SIP has been shown to be sensitive to change and emphasizes the impact of health upon activities and behavior, rather than emotional well-being (Bergner et al., 1981). The measure is scored using a weighted system whereby scale values are used as weights when summing the individual items to obtain the scale score for each of the 12 dimensions. A score from 0 (perfect health) to 100 (severe burden of morbidity) is produced. Two higher-order dimensions, consisting of scores for the physical and psychosocial domains are calculated and scored in a similar manner (Bergner et al., 1981). The SIP has been evaluated in ALS and findings support the validity of the measure (Damiano et al., 1999), However, studies have shown limitations in its usefulness when evaluating psychosocial well-being (Damiano et al., 1999; McGuire et al., 1996). A shorter version, the SIP/ALS-19 (McGuire et al., 1997), was developed specifically for use with ALS patients. The authors examined SIP subscales and derived a set of 19 clinically relevant items independently chosen by a panel of ALS experts. More extensive evaluation of the measure is required, however current studies have found the SIP/ALS-19 to be a valid measure

of physical function, but potentially not QOL (Bromberg et al., 2001; Jenkinson et al., 2011; Simmons et al., 2000).

The ALSSQOL-R is an ALS specific health profile (Felgoise et al., 2011). It was developed to be administered over time in clinic to guide assessments and interventions. It is a 50-item instrument composed of six specific domains: negative emotion, interaction with people and the environment, intimacy, religiosity, physical symptoms and bulbar function. Each domain is scored using a 0 (least desirable) to 10 (most desirable) point Likert scale. A single item QOL score, an average total QOL score and six domain scores are produced. Total scores range from 0 to 460 and are calculated by summing the scores for all questions and dividing by the number of questions answered. The same method can be used for calculating the subdomain scores. The ALSSQOL-R scale takes approximately 15-20 minutes to administer. Additionally, the reliability, validity and responsiveness of the measure has been evaluated and found to be adequate (Felgoise et al., 2011).

1.4.4 Preference-Based Measures

Preference-based measures (PBMs) (or utility measures) are similar to other outcome measures that assess HRQL by providing a structured way of including the patient's perspective when clinically evaluating the impairments, activities and social opportunities influenced by a disease and its treatment (Brazier, Ratcliffe, et al., 2017). However, this form of measure is derived from economic and decision theories that use preference-based methods to obtain the respondent's own overall value for dimensions their own HRQL (Brazier, Ratcliffe, et al., 2017; Guyatt, 1993). This is in contrast to non-preference approaches that mostly derive scores for dimensions from the summing of responses, such as with previously mentioned instruments (Fitzpatrick et al., 1998).

PBMs relate health states to death whereby they elicit a single value of HRQL that is anchored from 0 (death) to 1.0 (full health). For this reason, PBMs have become the most popular method to calculate quality-adjusted life years (QALYs) (Brazier, Ratcliffe, et al., 2017). QALYs express health as a function of both length of life (mortality) and quality of life (morbidity). The number of QALYs relating to a health outcome is expressed as the value given to a health state, multiplied by the length of time spent in that state (Brazier, Ratcliffe, et al., 2017). This value is widely used in economic evaluation for cost-utility analysis and enables PBMs to be used to evaluate the clinical effectiveness and cost-effectiveness of interventions (Brazier, Ratcliffe, et al., 2017). Furthermore, the single index produced reflects both the health status and the value of that health status to the patient, which allows comparisons across interventions and conditions (Brazier, Ratcliffe, et al., 2017; Guyatt, 1993).

There are two components to PBMs: a descriptive system for describing health or its impact on HRQL, and a scoring algorithm for assigning values to each state described by the system. The health state descriptive system is composed of several domains that together describe the HRQL of individuals. The scoring of each domain is provided by an algorithm based on preferences or values whereby values for PBMs can be obtained in one of two ways, directly or indirectly. Direct PBMs use methods such as standard gamble (SG), time trade off (TTO), or VAS to derive values. Indirect PBMs use questionnaires that are generic or disease-specific whereby the values are obtained either from the general population or the patient's themselves (Brazier, Ratcliffe, et al., 2017).

1.4.4.1 Direct Preference-Based Measures

The SG method uses expected utility theory (EUT) as its theoretical basis (Salomon, 2014). Respondents are asked to consider a choice between an outcome and two alternatives, one

being better, the other worse than the outcome. For example, in alternative A, a person would live with a possible health problem with certainty, for the remainder of his or her life. Alternative B is usually characterized by a risky treatment whereby they either live in a state of optimal health with probability p, or immediate death, with probability (1-p). The objective is to determine the point at which a respondent is indifferent to the alternatives. This probability is then considered the valuation for the health problem of interest (Salomon, 2014). The SG is considered the gold standard in the field of economics since it considers the uncertainty of decisions, similar to those decisions made in healthcare (Salomon, 2014). The SG has been found to be feasible, acceptable, and reliable and good response and completion rates in health contexts (Brazier, Ratcliffe, et al., 2017). However, criticisms underlie the values generated by SG as they may not represent people's valuation of a given health state and may indeed incorporate other factors such as risk or loss aversion, death, and cognitive burden (Brazier, Ratcliffe, et al., 2017).

The TTO method was originally devised specifically for use in healthcare as a 'short-cut' method of obtaining values equivalent to those using SG (Torrance, 1987). Respondents are asked to trade off duration of life against health status. They are not given probabilities and instead are asked their preference, for several different durations, e.g. one-year, nine-months etc., for a shorter life in a state of perfect health versus a longer life in a state of impaired health. Like the SG, the objective is to determine a point at which a respondent is indecisive. This method is based on choice, as respondents are asked to choose between two alternatives of certainty. Thus, it is likely easier to understand than probabilities (Torrance, 1987). It is a reliable and practical method of health state valuation (Green et al., 2000). However, criticisms have been made concerning the applicability of the TTO in healthcare as it is a measure of certainty (Brazier,

Ratcliffe, et al., 2017). Others have argued that there could be a time preference bias as individuals tend to give greater value to years of life in the near future than to those in the distant future (Brazier, Ratcliffe, et al., 2017).

The VAS is a commonly used format of rating scale, with two well-defined end points: in healthcare from best imaginable health state to worst imaginable health state (or death), on which respondents are asked to indicate their judgments, values or feelings (Brazier, Ratcliffe, et al., 2017). It was first identified as a possible measure for health economic evaluation over three decades ago (Patrick et al., 1973). This measure is feasible, and demonstrates high response rates and high levels of completion (Drummond, 2007; Froberg & Kane, 1989). However, as this method does not involve risk or choice there are concerns over the ability of the instrument to reflect preferences on an interval scale (Bleichrodt & Johannesson, 1997). Subsequently, Dyer and Sarin (1982) argued in favor of the use of VAS for health economic evaluation but specified it should only be used indirectly in the calculation of QALYs; by mapping preferences from the VAS onto SG or TTO utility values. Criticisms have also arisen concerning VAS methods as they tend to be: 1) prone to context effects where the average rating for items is influenced by the level of other items being valued (Brazier, Ratcliffe, et al., 2017), and 2) susceptible to response spreading, whereby similar health states are placed some distance from one another and health states that are vastly different are placed very close to each other (Brazier, Ratcliffe, et al., 2017).

1.4.4.2 Indirect Preference-Based Measures

It has been historically recognized that values for health can be obtained from patients, caregivers, health professionals and the community. While the most common method for PBMs is obtaining health state values from the general population, it has been argued that values should

be obtained from participants (Menzel et al., 2002; Ubel et al., 2003). The question then is whether to obtain preferences based on hypothetical or lived experience (Brazier et al., 2005).

Obtaining values from the general population is argued to provide a coherent set of values from society that can be used for generic PBMs (Brazier, Ratcliffe, et al., 2017). Public funding can essentially be seen as public insurance and a way for the public to contribute to the development of the measure. Moreover, the general population has no vested interest in the health condition(s) the measure is being developed for; therefore, they will appreciate how health states compare to those of other patients. Conversely, using patient values for PBMs allows for the addition of experience as patients know their own health state better than others knowing their health state (Brazier et al., 2005). Additionally, it is the well-being of the patient that we are interested in, thus patients should contribute to the values obtained. Despite the advantages for both, there are disadvantages to recognize. The general population has little to no first-hand experience of the health states being valued, which is an advantage of using patient preferences. However, patients can adapt to their health states which may result in them assigning a better health value to an otherwise undesirable state. Furthermore, patients may be unwilling or unable to undertake complex and quite intrusive valuation tasks (Brazier, Ratcliffe, et al., 2017).

1.5 Generic Preference-Based Measures of HRQL

Existing PBMs used in ALS are generic. Currently, there are seven leading generic preference-based measures (GPBMs): the EuroQol 5 Dimension (EQ-5D) (Bansback et al., 2012); Short Form 6 Dimension (SF-6D) (Brazier et al., 2002; Brazier & Roberts, 2004); Health Utilities Index Mark 2 (HUI2) and 3 (HUI3) (Abel et al., 2017; Simmons, 2015); the Assessment of QOL-8 Dimension (AQoL-8D) (Richardson et al., 2011); 15 Dimension (15D) (Sintonen,

2001) and Quality of Well-Being (QWB) scale (Kaplan & Anderson, 1988). These measures are widely used, and as they are generic should be both relevant to all patient groups and provide a means of making comparisons between patient groups. GPBMs are often used in clinical trials and routine data collection whereby the valuation of the responses is performed using the provided scoring algorithm, making these measures easy to administer and feasible.

1.5.1 EuroQol 5 Dimension (EQ-5D)

The EQ-5D-3L (Brooks, 1996) was developed by the EuroQol Group in 1996. It is a well-established and widely used generic preference-based measure of HRQL that consists of two parts (Brauer et al., 2006; Brazier, Ara, et al., 2017). The first part (the descriptive system) assesses health in five domains: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. In the EQ-5D-3L (3 level) version, 3 response levels are available for each domain, scored from 1 (no problems) to 3 (extreme problems). The second part (the scoring system) consists of a five-digit health utility score that is formed according to the responses from the five domains. In order to produce a single index score, a TTO derived value set with general population preferences for Canada is used and can result in health utilities ranging from -0.306 (33333) for the worst possible health state to 1.00 (11111) for the best possible health state (Bansback et al., 2012).

The EQ-5D-5L (5 level) version was later developed to improve the sensitivity of the measure and reduce ceiling effects (Herdman et al., 2011). The measure maintains the five domains from the EQ-5D-3L, but expands from 3 to 5 response levels (no, some, moderate, severe, extreme problems) (Herdman et al., 2011). The EQ-5D-5L defines a total of 3125 health states (5^5) (Herdman et al., 2011), a substantial increase from the EQ-5D-3L with 243 health states (3^5) (Brooks, 1996), and has been translated into more than 170 languages world-wide

(Reenen & Janssen, 2019). Both the EQ-5D-3L and EQ-5D-5L are self-administered measures that can be completed in less than 5 minutes. For the EQ-5D-5L, each domain is scored from 1 to 5 and a utility value is derived from the five questions. In order to produce a single index score, a TTO value set with general population preferences was recently developed for Canada and utilizes a model resulting in potential health utilities ranging from -0.148 for the worst possible health state (55555) to 0.949 (11111) for the best possible health state (Xie et al., 2016). In addition to the questionnaire, the EQ-5D measures contain a visual analogue scale (VAS) of self-rated health, scored from 0 to 100 (Brooks, 1996; Reenen & Janssen, 2019). The scores from the VAS cannot be used directly as weights in QALY calculations as they do not produce a single index value; however, the scores can be used as a subjective assessment of self-perceived health (Brooks, 1996).

1.5.2 Short Form 6 Dimension (SF-6D)

The SF-6D was derived from both the SF-36 and SF-12 (Brazier et al., 2002; Brazier & Roberts, 2004) for health economic evaluation purposes. The SF-6D consists of six domains: physical functioning, role limitations (physical and emotional), bodily pain, vitality, social functioning and mental health. The number of items and response levels vary for each domain, from one to three items and four to six response levels. The SF-6D (SF-36) describes a total of 18,000 health states (Brazier et al., 2002) whereas the SF-6D (SF-12) describes a total of 7500 health states (Brazier & Roberts, 2004). A sample from the general population in the United Kingdom was used to produce the health state valuations using SG. An additive model using an orthogonal design was used to generate the health states (Brazier et al., 2002; Brazier & Roberts, 2004). More recently, a new algorithm, found to perform better in terms of predictive ability and removing the bias seen in the original model, was estimated using a non-parametric Bayesian

approach (Kharroubi et al., 2007). A single index score from this method is produced that ranges from 0.20 (worst health) to 1.00 (perfect health). Both versions of the SF-6D are similar in psychometric properties and demonstrate responsiveness to change. However, similar to the observed floor effects reported with the SF-36 and the SF-12 (Bindman et al., 1990), there are also floor effects for the SF-6D (Brazier et al., 2004; Longworth & Bryan, 2003; O'Brien et al., 2003).

1.5.3 Health Utilities Index Mark 2 (HUI2) and 3 (HUI3)

The Health Utilities Index (HUI) is a family of generic measures that consist of the HUI1, HUI2 and HUI3 (Horsman et al., 2003). They were developed to focus on fundamental core attributes of health status and the capacity of individuals to function within these attributes. As a GPBM, they can be administered to individuals with various conditions and generate a single index score that can be used for economic evaluation.

The HUI1 was established to evaluate outcomes for very-low birth-weight infants (Horsman et al., 2003; Torrance et al., 1982). From the core set of four attributes, the HUI2 was then developed to address the global morbidity burden of childhood cancer (Feeny et al., 1996). The HUI2 consists of seven domains: sensations, mobility, emotion, cognition, self-care, pain and fertility (Feeny et al., 1995). Each domain has three to five response levels and describes 24,000 unique health states. The HUI2 scoring system and health values were developed by using general population preferences from the community (n=194) and by incorporating their VAS values and SG utilities in a multiplicative model (Torrance et al., 1996). A single utility score is produced that can range from 0.03 (worst health) to 1.00 (perfect health) (Horsman et al., 2003; Torrance et al., 1996). However, floor effects are present for the HUI2 (Torrance et al., 1996).

The HUI3 was developed to address concerns about definitions in the HUI2; for use in clinical and general population studies; and for structural independence among the attributes (Horsman et al., 2003). It includes eight HRQL domains that focus on bodily functions: vision, speech, hearing, dexterity, ambulation, cognition, emotion and pain. Each domain has five to six levels of function and describes a total of 972,000 unique health states. The HUI3 scoring system and health values were developed similar to the HUI2, by using general population preferences from a community sample (n=504) in Ontario, Canada and incorporating their ratings from single-deficit states using VAS and SG in a multiplicative model (Feeny et al., 2002). A single utility score is produced that can range from -0.36 (worse than death) to 1.00 (perfect health) (Feeny et al., 2002; Horsman et al., 2003).

The HUI instruments are self-completed measures that can be completed in 5 to 10 minutes. They have been found to be valid and reliable (Costet et al., 1998; Le Galès et al., 2002; Wang & Chen, 1999) and responsive to changes in health status over time (Blanchard et al., 2003; Furlong et al., 2001). Both measures are available in multiple languages and have been widely used in clinical populations, including neurological conditions (Edwards et al., 2010; Fisk, 2005; Pohar & Allyson Jones, 2009).

1.5.4 Assessment of Quality of Life 8 Dimension (AQoL-8D)

The Assessment of Quality of Life (AQoL) multi-attribute utility (MAU) instruments (AQoL-4D,6D,7D,8D) were developed with the intention of combining psychometric and utility measurement; to provide a health status profile whilst incorporating the utility property of a single HRQL index (Hawthorne et al., 1999).

The AQoL-8D is an extension of two earlier instruments, the AQoL-4D (originally called 'AQoL') (Hawthorne et al., 1999) and AQOL-6D (Richardson et al., 2004). It was developed in

order to achieve greater sensitivity by revising the descriptive system and scoring algorithm of the AQoL-6D (Richardson et al., 2012). The measures differ in dimensions and items. The domains of illness, independent living, social relationships, physical senses and psychological well-being were included in the 15 item AQoL-4D (Hawthorne et al., 1999). For the 20-item AQoL-6D, pain and coping were then added as domains and the response levels for mental health and independent living were increased by one (Richardson et al., 2004). For the 26-item AQoL-7D, an explicit dimension for vision was added (Misajon et al., 2005).

The 35-item AQoL-8D is the most recent measure developed (Richardson et al., 2011). It consists of 8 dimensions: independent living, pain, senses, mental health, happiness, coping, relationships and self-worth. Three of these are related to the physical and the remaining five to the psychosocial. Each domain has two to ten items with four to six response levels for each resulting in a measure that describes a total of 2.4×10^{23} health states (Richardson et al., 2014). The scoring system and health utility values were developed by using a sample from the general health population (n=347) and mental health population (n=323). A multiplicative model was used to combine items into dimensions from VAS and TTO derived (and predicted) values (Richardson et al., 2011). A single index score is produced that can range from 0.17 (worst health) to 1.00 (perfect health) (Richardson et al., 2011). The AQoL-8D has been found to be a reliable and valid instrument (Hawthorne et al., 2001; Richardson et al., 2014). The AQoL-8D has been administered in various clinical populations and is responsive to changes in health status (Ahmad et al., 2020; Campbell et al., 2018; Dieng et al., 2018). More research is needed in the application of the measure in progressive conditions as one study (Grivell et al., 2018) found evidence that the AQoL-8D might not be responsive to change in HRQL in ALS.

1.5.5 15 Dimension (15D)

The 15 Dimension (15D) is a self-reported, GPBM of HRQL for individuals (Sintonen, 2001). It was initially developed as a 12-dimension measure in 1981 to combine the advantages of a profile and a preference-based measure (Sintonen, 1981). In 1986, a 15-dimension revised version (15D.1) of that measure (Sintonen & Pekurinen, 1989) was established, based on feedback from the physicians. The measure underwent one final revision of its descriptive system in 1992 (15D.2) in order to increase its sensitivity (Sintonen, 1994).

The 15D (15D.2) includes the following 15 domains: breathing, mental function, speech (communication), vision, mobility, usual activities, vitality, hearing, eating, elimination, sleeping, distress, discomfort and symptoms, sexual activity and depression (Sintonen, 1994). Each domain/item has four or five response levels, resulting in a measure that describes billions of health states. The scoring and valuation system of the 15D produces a single index score from 0 (death) to 1 (perfect health) and was derived from general population preferences in a similar method to a VAS however with ratio properties (Sintonen, 1995). A simple additive formula was then used to estimate the health states values. The 15D has been administered in various chronic and neurological conditions and has been used for economic evaluation purposes (Haapaniemi, 2004; Lunde, 2013; Saarni et al., 2006; Stavem et al., 2001).

1.5.6 Quality of Well Being (QWB) Scale

The Quality of Well-Being (QWB) scale is the oldest of the QALY instruments. It was developed as part of the General Health Policy Model to inform resource allocation in health services and to include mortality and morbidity as specific components of HRQL (Kaplan & Anderson, 1988). The QWB scale combines three scales of functioning (mobility, physical activity and social activity) with a measure of symptoms and problems to produce a single index score with anchors at 0 (death) and 1 (asymptomatic full function) (Kaplan & Anderson, 1988).

The measure is unique in incorporating both a functioning component and a strong symptom component in the items included in the measure. The domains consist of three levels each resulting in a total of 46 functional levels, in addition to a list of 27 symptoms with two response levels each. The QWB scale describes a total of 945 health states and takes approximately 7 to 20 minutes to administer (Bombardier & Raboud, 1991; Kaplan & Anderson, 1988), with one to two weeks of training for the interviewers. Preference weights and valuations were derived for the QWB scale from a community sample (n=866) in San Diego. Each respondent valued a random sample of 42 health states using a VAS where zero corresponded to death. Preference weights were calculated using an additive model, and range from 0.33 (worse health) to 1.00 (perfect health) (Seiber et al., 2008).

In response to criticisms of the QWB scale, a self-administered version of the scale was developed (QWB-SA). The QWB-SA scale is a comprehensive measure encompassing 58 symptoms (mental, acute physical and chronic) compared to the 26 symptoms of the QWB scale (Seiber et al., 2008). There are five sections to the measure. The first section asks about the presence of 19 chronic, 25 acute physical, and 14 mental health symptoms or problems. The remaining sections are similar to the QWB scale and include assessments of self-care, mobility, physical activity, performance of physical functioning, and performance of usual activity (Kaplan et al., 1997). The QWB-SA takes approximately 14 minutes to complete (Andresen et al., 1998). Preference weights were derived from a community sample in San Diego (n=430) and each subject completed a randomly selected subsample of 12 items using a VAS. An adaptation of an additive model was used to calculate the weights for all items and then weights for the three domains were calculated by subtraction (Seiber et al., 2008). Preference weights derived range from 0.09 (worse health) to 1.00 (perfect health) (Seiber et al., 2008). Both the QWB and

QWB-SA scale are comparable in scores (Kaplan et al., 1997) and demonstrate reliability, validity and responsiveness in various conditions (Andresen et al., 1998; Kaplan et al., 1989, 1998; Kerner et al., 1998).

1.6 Psychometric Properties of HRQL Measures

Determining which HRQL measure is the most effective and best to use is a question of importance for researchers and health care professionals. Without psychometric testing and valid outcome measures, researchers would be unable to determine whether an intervention has had an impact. Moreover, without valid and reliable interpretations of scores obtained, inaccurate decisions could be made regarding which GPBM is the most valid and reliable. For instance, studies have ascertained the use of GPBMs in various health conditions (Brazier, Ara, et al., 2017), such as stroke (Simon Pickard et al., 2005). These measures have established estimates of psychometric reliability and validity and can be used to make comparisons across diseases and interventions in these populations. However, if these properties are not tested, scores from instruments may overestimate or underestimate the HRQL of individuals (Brazier, Ratcliffe, et al., 2017). Therefore, a set criterion, namely the psychometric properties of reliability, validity and responsiveness should be used in practice to ensure that the values obtained by the scoring system are valid for interpretation and utilization by healthcare professionals, researchers and policy makers.

1.6.1 Reliability

Reliability is concerned with the reproducibility of an instrument. It is defined as the extent to which a measure provides the same results in repeated trials or from different raters (Mokkink et al., 2010). It is essential to establish that any changes observed are due to the

intervention and not the instrument itself. Furthermore, reliability is a necessary condition for validity, as a measure must be reliable to be valid.

There are four types of reliability: test-retest reliability, interrater reliability, intra-rater reliability and internal consistency. Test-retest reliability is defined as the extent to which scores of a measure have not changed over time, provided the characteristics being measured do not change (Mokkink et al., 2010; Streiner et al., 2015). Interrater and intra-rater reliability refers to the degree to which different raters' or observers' scores on the same measure agree; whether by different raters on the same occasion or by the same raters on different occasions, respectively (Mokkink et al., 2010). Internal consistency is often measured by Cronbach's alpha and refers to the relationship among items administered at the same point in time (Mokkink et al., 2010).

1.6.2 Validity

The validity of a measure is concerned with the extent to which an instrument truly measures what it purports to measure; that is if the instrument truly captures what it says it measures, i.e. the construct of interest. There are several different ways of capturing the validity of a measure in relation to a specific purpose or set of purposes: face and content validity, construct validity (i.e. convergent and discriminative (known-groups) validity), and criterion and predictive validity (de Vet, 2011).

Face validity is the degree to which an instrument looks as though it is an adequate reflection of the construct to be measured (Mokkink et al., 2010). It is a subjective assessment that is most often evaluated during the development of a measure rather than during use. If face validity is not present, it is a very strong argument for not testing an instrument. Content validity is defined as the degree to which the content of an instrument is an adequate reflection of the construct of interest (Mokkink et al., 2010). Items should be relevant and comprehensive for the

construct of interest. It is assessed during the development of an instrument or by users of the instrument (de Vet, 2011). Content validity is important to evaluate as a measure needs to capture the full range of possible presentations of the construct of interest. This can be affected by floor and ceilings effects that if present, do not allow a measure to fully capture the range of patient experiences (de Vet, 2011).

Construct validity is applicable in situations where there is no gold standard. It refers to how well scores of the instrument provide expected scores and is separated into convergent validity and discriminative (known-groups validity). Convergent validity is defined as the degree to which scores of two measurement instruments relate when measuring a similar construct of interest (Chin & Yao, 2014; Mokkink et al., 2010). One would expect similar scores or a higher correlation when measuring instruments that are supposed to be similar in their assessment of the construct. Discriminative validity, on the other hand, is defined as the degree to which an instrument is able to discriminate between two groups that differ on the construct being measured (Davidson, 2014). It is the extent to which scores on a measure do not share a relationship with scores obtained on a theoretically unrelated measure.

Last, criterion validity is applicable in situations where there is a gold standard for the construct to be measured. In general, in the field of HRQL, there is no gold standard and thus criterion validity is of limited use aside from predictive validity (de Vet, 2011; Fitzpatrick et al., 1998) . Predictive validity is defined as the extent to which measurement instrument scores is an adequate reflection of a gold standard for the construct of interest in the future (de Vet, 2011). This type of validity is concerned with whether the instrument predicts future behavior from scores on a measure.

1.6.3 Responsiveness

Responsiveness is considered an important aspect of validity. It is defined as the ability of an instrument to detect change over time in the construct of interest (Mokkink et al., 2010). The ability to detect change is essential for an instrument to demonstrate since it indicates a measure of sensitivity to change in the instrument and the validity of a change score. A longitudinal study is required to assess responsiveness as at least two measurements should be taken in order to calculate change scores (de Vet, 2011).

1.6.4 Interpretability

The interpretability of a measure is not considered a psychometric property however, it is an important characteristic of a measurement instrument. It is defined as the degree to which one can assign qualitative meaning (clinical or commonly understood connotations) to an instrument's quantitative scores or change in scores (Mokkink et al., 2010). Interpretation generally involves an idea of the magnitude of differences or changes that are clinically important. In addition, floor and ceiling effects must also be considered. Floor and ceiling effects are defined as the percentage of the sample obtaining scores at the lower and upper ends of the scale, respectively (de Vet, 2011). These effects can alter the responsiveness of an instrument whereby an overestimation or underestimation of the construct can result in a measure that is not able to fully capture the effects of the condition on the construct of interest (de Vet, 2011). Floor and ceiling effects are deemed significant when percentage values greater than 15% are seen (McHorney & Tarlov, 1995).

1.7 Rationale

This thesis examines the psychometric properties of GPBMs of HRQL in ALS. The reliability and validity of these measures have not yet been evaluated in ALS, resulting in an

important gap in the literature that needs to be addressed. This is especially concerning as ALS affects all aspects of an individual's life, including HRQL, and treatment is focused on addressing this construct. Current PBMs in ALS are generic and can be used for economic evaluation. However, as GPBMs are not developed for use in ALS, it is of utmost importance to assess their psychometric properties in this population and determine whether the content of the instruments accurately capture areas of HRQL impacted by ALS. This thesis will report the results from a systematic review with the primary objective of synthesizing the psychometric properties of GPBMs in ALS. Building on the information found, a second study will report on the content validity of GPBMs in ALS and the convergent validity of the EQ-5D-5L against the PGI.

References

Abel, H., Kephart, G., Packer, T., & Warner, G. (2017). Discordance in Utility Measurement in Persons with Neurological Conditions: A Comparison of the SF-6D and the HUI3. *Value in Health*, *20*(8), 1157–1165. https://doi.org/10.1016/j.jval.2017.04.008

Ahmad, H., van der Mei, I., Taylor, B. V., Campbell, J. A., & Palmer, A. J. (2020). Measuring the health-related quality of life in Australians with multiple sclerosis using the assessment of quality of life-8-dimension (AQoL-8D) multi-attribute utility instrument. *Multiple Sclerosis and Related Disorders*, *44*, 102358. https://doi.org/10.1016/j.msard.2020.102358

Andresen, E. M., Rothenberg, B. M., & Kaplan, R. M. (1998). Performance of a self-administered mailed version of the Quality of Well-Being (QWB-SA) questionnaire among older adults. *Medical Care*, *36*(9), 1349–1360. https://doi.org/10.1097/00005650-199809000-00007

Bakas, T., McLennon, S. M., Carpenter, J. S., Buelow, J. M., Otte, J. L., Hanna, K. M., Ellett, M. L., Hadler, K. A., & Welch, J. L. (2012). Systematic review of health-related quality of life models. *Health and Quality of Life Outcomes*, *10*(1), 134. https://doi.org/10.1186/1477-7525-10-134

Bansback, N., Tsuchiya, A., Brazier, J., & Anis, A. (2012). Canadian Valuation of EQ-5D Health States: Preliminary Value Set and Considerations for Future Valuation Studies. *PLoS ONE*, *7*(2), e31115. https://doi.org/10.1371/journal.pone.0031115

Bäumer, D., Talbot, K., & Turner, M. R. (2014). Advances in motor neurone disease. *Journal of the Royal Society of Medicine*, *107*(1), 14–21. https://doi.org/10.1177/0141076813511451

Bede, P., Oliver, D., Stodart, J., van den Berg, L., Simmons, Z., O Brannagain, D., Borasio, G. D., & Hardiman, O. (2011). Palliative care in amyotrophic lateral sclerosis: A review of

current international guidelines and initiatives. *Journal of Neurology, Neurosurgery & Psychiatry*, *82*(4), 413–418. https://doi.org/10.1136/jnnp.2010.232637

Bergner, M. (1989). Quality of Life, Health Status, and Clinical Research. *Medical Care*, *27*(3), S148–S156. JSTOR. www.jstor.org/stable/3765660

Bergner, M., Bobbitt, R. A., Carter, W. B., & Gilson, B. S. (1981). The Sickness Impact Profile: Development and final revision of a health status measure. *Medical Care*, *19*(8), 787–805. https://doi.org/10.1097/00005650-198108000-00001

Bindman, A. B., Keane, D., & Lurie, N. (1990). Measuring Health Changes Among Severely III Patients: The Floor Phenomenon. *Medical Care*, *28*(12), 1142–1152. https://doi.org/10.1097/00005650-199012000-00003

Blanchard, C., Feeny, D., Mahon, J. L., Bourne, R., Rorabeck, C., Stitt, L., & Webster-Bogaert, S. (2003). Is the health utilities index responsive in total hip arthroplasty patients? *Journal of Clinical Epidemiology*, *56*(11), 1046–1054. https://doi.org/10.1016/S0895-4356(03)00203-8

Bleichrodt, H., & Johannesson, M. (1997). An Experimental Test of a Theoretical Foundation for Rating-scale Valuations. *Medical Decision Making*, *17*(2), 208–216. https://doi.org/10.1177/0272989X9701700212

Bombardier, C., & Raboud, J. (1991). A comparison of health-related quality-of-life measures for rheumatoid arthritis research. *Controlled Clinical Trials*, *12*(4), S243–S256. https://doi.org/10.1016/S0197-2456(05)80028-5

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (Third edition). Routledge, Taylor and Francis Group.

Brauer, C. A., Rosen, A. B., Greenberg, D., & Neumann, P. J. (2006). Trends in the Measurement of Health Utilities in Published Cost-Utility Analyses. *Value in Health*, *9*(4), 213–218. https://doi.org/10.1111/j.1524-4733.2006.00116.x

Brazier, J., Akehurst, R., Brennan, A., Dolan, P., Claxton, K., McCabe, C., Sculpher, M., & Tsuchyia, A. (2005). Should Patients Have a Greater Role in Valuing Health States?: *Applied Health Economics and Health Policy*, *4*(4), 201–208. https://doi.org/10.2165/00148365-200504040-00002

Brazier, J., Ara, R., Rowen, D., & Chevrou-Severac, H. (2017). A Review of Generic Preference-Based Measures for Use in Cost-Effectiveness Models. *PharmacoEconomics*, *35*(S1), 21–31. https://doi.org/10.1007/s40273-017-0545-x

Brazier, J., Harper, R., Munro, J., Walters, S. J., & Snaith, M. L. (1999). Generic and condition-specific outcome measures for people with osteoarthritis of the knee. *Rheumatology*, *38*(9), 870–877. https://doi.org/10.1093/rheumatology/38.9.870

Brazier, J., Ratcliffe, J., Salomon, J. A., & Tsuchiya, A. (2017). *Measuring and valuing health benefits for economic evaluation* (Second edition). Oxford University Press.

Brazier, J., & Roberts, J. (2004). The Estimation of a Preference-Based Measure of Health From the SF-12. *Medical Care*, *42*(9), 851–859. https://doi.org/10.1097/01.mlr.0000135827.18610.0d

Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*, *21*(2), 271–292. https://doi.org/10.1016/S0167-6296(01)00130-8

Brazier, J., Roberts, J., Tsuchiya, A., & Busschbach, J. (2004). A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics*, *13*(9), 873–884. https://doi.org/10.1002/hec.866

Bromberg, M. B., Anderson, F., Davidson, M., & Miller, R. G. (2001). Assessing health status quality of life in ALS: Comparison of the SIP/ALS-19 with the ALS Functional Rating Scale and the Short Form-12 Health Survey. *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders*, *2*(1), 31–37. https://doi.org/10.1080/146608201300079391

Brooks, B. R. (1994). El escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. *Journal of the Neurological Sciences*, *124*, 96–107. https://doi.org/10.1016/0022-510X(94)90191-0

Brooks, B. R., Miller, R. G., Swash, M., & Munsat, T. L. (2000). El Escorial revisited: Revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders*, *1*(5), 293–299. https://doi.org/10.1080/146608200300079536

Brooks, B. R., Ringel, S., England, J., Brinkmann, J., Pestronk, A., Florence, J., Mitsumoto, H., Szirony, K., Wittes, J., Charatan, M., Stambler, N., & Cedarbaum, J. M. (1996). *The ALS functional rating scale: Assessment of activities of daily living in patients with amyotrophic lateral sclerosis*. *53*, 141–147.

Brooks, R. (1996). EuroQol: The current state of play. *Health Policy*, *37*(1), 53–72. https://doi.org/10.1016/0168-8510(96)00822-6

Calman, K. C. (1984). Quality of life in cancer patients—An hypothesis. *Journal of Medical Ethics*, *10*(3), 124–127. https://doi.org/10.1136/jme.10.3.124

Campbell, J. A., Hensher, M., Neil, A., Venn, A., Otahal, P., Wilkinson, S., & Palmer, A. J. (2018). An Exploratory Study: A Head-to-Head Comparison of the EQ-5D-5L and AQoL-8D for Long-Term Publicly Waitlisted Bariatric Surgery Patients Before and 3 Months After Bariatric Surgery. *PharmacoEconomics - Open*, *2*(4), 443–458. https://doi.org/10.1007/s41669-017-0060-1

Canadian Agency for Drugs and Technologies in Health. (2019). *CADTH Canadian Drug Expert Committee Recommendation: Edaravone (Radicava — Mitsubishi Tanabe Pharma Corporation): Indication: For the treatment of amyotrophic lateral sclerosis*. Canadian Agency for Drugs and Technologies in Health. http://www.ncbi.nlm.nih.gov/books/NBK542405/

Cedarbaum, J. M., & Stambler, N. (1997). Performance of the Amyotrophic Lateral Sclerosis Functional Rating Scale (ALSFRS) in multicenter clinical trials. *Journal of the Neurological Sciences*, *152*, s1–s9. https://doi.org/10.1016/S0022-510X(97)00237-2

Cedarbaum, J. M., Stambler, N., Malta, E., Fuller, C., Hilt, D., Thurmond, B., & Nakanishi, A. (1999). The ALSFRS-R: A revised ALS functional rating scale that incorporates assessments of respiratory function. *Journal of the Neurological Sciences*, *169*(1–2), 13–21. https://doi.org/10.1016/S0022-510X(99)00210-5

Charcot, J.-M., & Joffroy, A. (1869). *Deux cas d'atrophie musculaire progressive: Avec lésions de la substance grise et des faisceaux antérolatéraux de la moelle épinière*. *2*(7), 44–754. /z-wcorg/.

Chin, C.-L., & Yao, G. (2014). Convergent Validity. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 1275–1276). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_573

Cieza, A., & Stucki, G. (2008). The International Classification of Functioning Disability and Health: Its development process and content validity. *European Journal of Physical and Rehabilitation Medicine*, *44*(3), 303–313.

Costet, N., Le Galès, C., Buron, C., Kinkor, F., Mesbah, M., Chwalow, J., & Slama, G. (1998). French cross-cultural adapatation of the health utlities index Mark 2 (HUI2) and 3 (HUI3) classification systemats. Clinical and economic working groups. *Quality of Life Research*, *7*(3), 245–256. https://doi.org/10.1023/A:1008830115246

Cummins, R. A. (2005). Moving from the quality of life concept to a theory. *Journal of Intellectual Disability Research*, *49*(10), 699–706. https://doi.org/10.1111/j.1365-2788.2005.00738.x

Damiano, A. M., Patrick, D. L., Guzman, G. I., Gawel, M. J., Gelinas, D. F., Natter, H. M., & Ingalls, K. K. (1999). Measurement of Health-Related Quality of Life in Patients With Amyotrophic Lateral Sclerosis in Clinical Trials of New Therapies: *Medical Care*, *37*(1), 15–26. https://doi.org/10.1097/00005650-199901000-00004

Davidson, M. (2014). Known-Groups Validity. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 3481–3482). Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_1581

de Vet, H. C. W. (Ed.). (2011). *Measurement in medicine: A practical guide*. Cambridge University Press.

Dieng, M., Kasparian, N. A., Cust, A. E., Costa, D. S. J., Tran, A., Butow, P. N., Menzies, S. W., Mann, G. J., & Morton, R. L. (2018). Sensitivity of Preference-Based Quality-of-Life Measures for Economic Evaluations in Early-Stage Melanoma. *JAMA Dermatology*, *154*(1), 52–59. https://doi.org/10.1001/jamadermatol.2017.4701

Drummond, M. F. (Ed.). (2007). *Methods for the economic evaluation of health care programmes* (3. ed., reprint). Oxford Univ. Press.

Dyer, J. S., & Sarin, R. K. (1982). Relative Risk Aversion. *Management Science*, *28*(8), 875–886. https://doi.org/10.1287/mnsc.28.8.875

Edwards, J. D., Koehoorn, M., Boyd, L. A., & Levy, A. R. (2010). Is Health-Related Quality of Life Improving After Stroke?: A Comparison of Health Utilities Indices Among Canadians With Stroke Between 1996 and 2005. *Stroke*, *41*(5), 996–1000. https://doi.org/10.1161/STROKEAHA.109.576678

Fayers, P. M., & Machin, D. (2016). *Quality of life: The assessment, analysis, and reporting of patient-reported outcomes* (Third edition). John Wiley & Sons Inc.

Feeny, D., Furlong, W., Boyle, M., & Torrance, G. W. (1995). Multi-Attribute Health Status Classification Systems: Health Utilities Index. *PharmacoEconomics*, *7*(6), 490–502. https://doi.org/10.2165/00019053-199507060-00004

Feeny, D., Furlong, W., Torrance, G. W., Goldsmith, C. H., Zhu, Z., Depauw, S., Denton, M., & Boyle, M. (2002). Multiattribute and Single-Attribute Utility Functions for the Health Utilities Index Mark 3 System: *Medical Care*, *40*(2), 113–128. https://doi.org/10.1097/00005650-200202000-00006

Feeny, D., Torrance, G. W., & Furlong, W. (1996). Health Utilities Index. In B. Spilker (Ed.), *Quality of Life and Pharmacoeconomics in Clinical Trials* (2nd ed., pp. 239–252). Lippincott-Raven Press.

Felce, D., & Perry, J. (1995). Quality of life: Its definition and measurement. *Research in Developmental Disabilities*, *16*(1), 51–74. https://doi.org/10.1016/0891-4222(94)00028-8

Felgoise, S. H., Walsh, S. M., Stephens, H. E., Brothers, A., & Simmons, Z. (2011). *The ALS Specific Quality of Life-Revised (ALSSQOL-R) User's Guide*.

Ferrans, C. E., Zerwic, J. J., Wilbur, J. E., & Larson, J. L. (2005). Conceptual Model of Health-Related Quality of Life. *Journal of Nursing Scholarship*, *37*(4), 336–342. https://doi.org/10.1111/j.1547-5069.2005.00058.x

Fisk, J. D. (2005). A comparison of health utility measures for the evaluation of multiple sclerosis treatments. *Journal of Neurology, Neurosurgery & Psychiatry*, *76*(1), 58–63. https://doi.org/10.1136/jnnp.2003.017897

Fitzpatrick, R., Davey, C., Buxton, M. J., & Jones, D. R. (1998). Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment (Winchester, England)*, *2*(14), i–iv, 1–74.

Franchignoni, F., Mandrioli, J., Giordano, A., Ferro, S., & ERRALS Group. (2015). A further Rasch study confirms that ALSFRS-R does not conform to fundamental measurement requirements. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *16*(5–6), 331–337. https://doi.org/10.3109/21678421.2015.1026829

Franchignoni, F., Mora, G., Giordano, A., Volanti, P., & Chiò, A. (2013). Evidence of multidimensionality in the ALSFRS-R Scale: A critical appraisal on its measurement properties using Rasch analysis. *Journal of Neurology, Neurosurgery & Psychiatry*, *84*(12), 1340–1345. https://doi.org/10.1136/jnnp-2012-304701

Froberg, D. G., & Kane, R. L. (1989). Methodology for measuring health-state preferences—II: Scaling methods. *Journal of Clinical Epidemiology*, *42*(5), 459–471. https://doi.org/10.1016/0895-4356(89)90136-4

Furlong, W., Feeny, D., Torrance, G. W., & Barr, R. D. (2001). The Health Utilities Index (HUI®) system for assessing health-related quality of life in clinical studies. *Annals of Medicine*, *33*(5), 375–384. https://doi.org/10.3109/07853890109002092

Gordon, P. H. (2013). Amyotrophic Lateral Sclerosis: An update for 2013 Clinical Features, Pathophysiology, Management and Therapeutic Trials. *Aging and Disease*, *04*(05), 295–310. https://doi.org/10.14336/AD.2013.0400295

Green, C., Brazier, J., & Deverill, M. (2000). Valuing Health-Related Quality of Life: A Review of Health State Valuation Techniques. *PharmacoEconomics*, *17*(2), 151–165. https://doi.org/10.2165/00019053-200017020-00004

Grivell, N., Aiyappan, V., Catcheside, P., Keighley-James, G., Schultz, D., Glaetzer, K., Allcroft, P., Antic, N. A., & McEvoy, D. (2018). *Quality of life assessments in motor neurone disease patients with non-invasive ventilation using disease specific, treatment specific and generic tools*. *197*.

Guyatt, G. H. (1993). Measuring Health-Related Quality of Life. *Annals of Internal Medicine*, *118*(8), 622. https://doi.org/10.7326/0003-4819-118-8-199304150-00009

Haapaniemi, T. H. (2004). The generic 15D instrument is valid and feasible for measuring health related quality of life in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *75*(7), 976–983. https://doi.org/10.1136/jnnp.2003.015693

Haas, B. K. (1999). A Multidisciplinary Concept Analysis of Quality of Life. *Western Journal of Nursing Research*, *21*(6), 728–742. https://doi.org/10.1177/01939459922044153

Hardiman, O., van den Berg, L. H., & Kiernan, M. C. (2011). Clinical diagnosis and management of amyotrophic lateral sclerosis. *Nature Reviews Neurology*, *7*(11), 639–649. https://doi.org/10.1038/nrneurol.2011.153

Harper, R., Brazier, J., Waterhouse, J. C., Walters, S. J., Jones, N. M., & Howard, P. (1997). Comparison of outcome measures for patients with chronic obstructive pulmonary disease (COPD) in an outpatient setting. *Thorax*, *52*(10), 879–887. https://doi.org/10.1136/thx.52.10.879

Hawthorne, G., Richardson, J., & Day, N. A. (2001). A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. *Annals of Medicine*, *33*(5), 358–370. https://doi.org/10.3109/07853890109002090

Hawthorne, G., Richardson, J., & Osborne, R. (1999). The Assessment of Quality of Life (AQoL) Instrument: A Psychometric Measure of Health-Related Quality of Life. *Quality of Life Research*, *8*(3), 209–224. JSTOR. www.jstor.org/stable/4035815

Herdman, M., Gudex, C., Lloyd, A., Janssen, Mf., Kind, P., Parkin, D., Bonsel, G., & Badia, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*, *20*(10), 1727–1736. https://doi.org/10.1007/s11136-011-9903-x

Hickey, A. M., Bury, G., O'Boyle, C. A., Bradley, F., O'Kelly, F. D., & Shannon, W. (1996). A new short form individual quality of life measure (SEIQoL-DW): Application in a cohort of individuals with HIV/AIDS. *BMJ*, *313*(7048), 29–33. https://doi.org/10.1136/bmj.313.7048.29

Horsman, J., Furlong, W., Feeny, D., & Torrance, G. W. (2003). The Health Utilities Index (HUI): Concepts, measurement properties and applications. *Health and Quality of Life Outcomes*, *1*, 54. https://doi.org/10.1186/1477-7525-1-54

Ingre, C., Roos, P. M., Piehl, F., Kamel, F., & Fang, F. (2015). Risk factors for amyotrophic lateral sclerosis. *Clinical Epidemiology*, *7*, 181–193. https://doi.org/10.2147/CLEP.S37505

Jenkinson, C., Hobart, J., Chandola, T., Fitzpatrick, R., Peto, V., Swash, M., & The ALS-HPS Steering Group. (2002). Use of the short form health survey (SF-36) in patients with amyotrophic lateral sclerosis: Tests of data quality, score reliability, response rate and scaling assumptions. *Journal of Neurology*, *249*(2), 178–183. https://doi.org/10.1007/PL00007861

Jenkinson, C., Layte, R., Jenkinson, D., Lawrence, K., Petersen, S., Paice, C., & Stradling, J. (1997). *A shorter form health survey: Can the SF-12 replicate results from the SF-36 in longitudinal studies? 19*(2), 179–186.

Jenkinson, C., Peters, M., & Bromberg, M. B. (Eds.). (2011). Patient-reported outcome measurement in motor neuron disease/amyotrophic lateral sclerosis—The ALSAQ-40 and ALSAQ-5. In *Quality of life measurement in neurodegenerative and related conditions* (pp. 41–49). Cambridge University Press.

Joyce, C. R. B., McGee, H. M., & OḂoyle, C. A. (1999). *Individual quality of life: Approaches to conceptualisation and assessment*. Harwood Academic Publishers. http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=526533

Kaplan, R. M. (1985). Quality of life measurement. In P. Karoly (Ed.), *Measurement Strategies in Health Psychology*. John Wiley.

Kaplan, R. M., & Anderson, J. P. (1988). A general health policy model: Update and applications. *Health Services Research*, *23*(2), 203–235.

Kaplan, R. M., Ganiats, T., Sieber, W., & Anderson, J. P. (1998). The quality of well-being scale: Critical similarities and differences with SF-36. *International Journal for Quality in Health Care*, *10*(6), 509–520. https://doi.org/10.1093/intqhc/10.6.509

Kaplan, R. M., Wu, A. W., Mathews, W. C., Kozin, F., & Orenstein, D. (1989). The Quality of Well-Being Scale: Applications in AIDS, Cystic Fibrosis, and Arthritis. *Medical Care*, *27*(Supplement), S27–S43. https://doi.org/10.1097/00005650-198903001-00003

Kaplan, R. M., Sieber, W. J., & Ganiats, T. G. (1997). The quality of well-being scale: Comparison of the interviewer-administered version with a self-administered questionnaire. *Psychology & Health*, *12*(6), 783–791. https://doi.org/10.1080/08870449708406739

Karanevich, A. G., Weisbrod, L. J., Jawdat, O., Barohn, R. J., Gajewski, B. J., He, J., & Statland, J. M. (2018). Using automated electronic medical record data extraction to model ALS survival and progression. *BMC Neurology*, *18*(1), 205. https://doi.org/10.1186/s12883-018-1208-z

Karimi, M., & Brazier, J. (2016). Health, Health-Related Quality of Life, and Quality of Life: What is the Difference? *PharmacoEconomics*, *34*(7), 645–649. https://doi.org/10.1007/s40273-016-0389-9

Kerner, D. N., Patterson, T. L., Grant, I., & Kaplan, R. M. (1998). Validity of the Quality of Well-Being Scale for Patients with Alzheimer's Disease. *Journal of Aging and Health*, *10*(1), 44–61. https://doi.org/10.1177/089826439801000103

Kharroubi, S. A., Brazier, J., Roberts, J., & O'Hagan, A. (2007). Modelling SF-6D health state preference data using a nonparametric Bayesian method. *Journal of Health Economics*, *26*(3), 597–612. https://doi.org/10.1016/j.jhealeco.2006.09.002

Kiebert G.M., Green C., Murphy C., Mitchell J.D., O'Brien M., Burrell A., & Leigh P.N. (2001). Patients' health-related quality of life and utilities associated with different stages of amyotrophic lateral sclerosis. *Journal of the Neurological Sciences*, *191*(1–2), 87–93. https://doi.org/10.1016/S0022-510X%2801%2900616-5

Kiernan, M. C., Vucic, S., Cheah, B. C., Turner, M. R., Eisen, A., Hardiman, O., Burrell, J. R., & Zoing, M. C. (2011). Amyotrophic lateral sclerosis. *Lancet (London, England)*, *377*(9769), 942–955. https://doi.org/10.1016/S0140-6736(10)61156-7

Kuspinar, A., Mate, K., Lafontaine, A.-L., & Mayo, N. (2019). Evaluating the content validity of generic preference-based measures for use in Parkinson's disease. *Parkinsonism & Related Disorders*, *62*, 112–116. https://doi.org/10.1016/j.parkreldis.2019.01.014

Kuspinar, A., & Mayo, N. E. (2013). Do generic utility measures capture what is important to the quality of life of people with multiple sclerosis? *Health and Quality of Life Outcomes*, *11*(1), 71. https://doi.org/10.1186/1477-7525-11-71

Kuyken, W., & Group, T. (1995). *The World Health Organization Quality of Life assessment (WHOQOL): Position paper from the World Health Organization*. *41*, 1403–1409.

Le Galès, C., Buron, C., Costet, N., Rosman, S., & Slama, Pr. G. (2002). Development of a preference-weight health status classification system in france: The Health Utilities Index 3. *Health Care Management Science*, *5*(1), 41–51. https://doi.org/10.1023/A:1013201102918

Leighton, D. J., Newton, J., Colville, S., Davenport, R., Gorrie, G., Morrison, I., Swingler, R., Chandran, S., & Pal, S. (2019). Changing epidemiology of motor neurone disease in Scotland. *Journal of Neurology*, *266*(4), 817–825. https://doi.org/10.1007/s00415-019-09190-7

Longinetti, E., & Fang, F. (2019). Epidemiology of amyotrophic lateral sclerosis: An update of recent literature. *Current Opinion in Neurology*, *32*(5), 771–776. https://doi.org/10.1097/WCO.0000000000000730

Longworth, L., & Bryan, S. (2003). An empirical comparison of EQ-5D and SF-6D in liver transplant patients. *Health Economics*, *12*(12), 1061–1067. https://doi.org/10.1002/hec.787

Lou, J.-S., Reeves, A., Benice, T., & Sexton, G. (2003). Fatigue and depression are associated with poor quality of life in ALS. *Neurology*, *60*(1), 122–123. https://doi.org/10.1212/01.WNL.0000042781.22278.0A

Ludolph, A., Drory, V., Hardiman, O., Nakano, I., Ravits, J., Robberecht, W., Shefner, J., & for The WFN Research Group On ALS/MND. (2015). A revision of the El Escorial criteria—2015. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *16*(5–6), 291–292. https://doi.org/10.3109/21678421.2015.1049183

Lunde, L. (2013). Can EQ-5D and 15D be used interchangeably in economic evaluations? Assessing quality of life in post-stroke patients. *The European Journal of Health Economics*, *14*(3), 539–550. https://doi.org/10.1007/s10198-012-0402-y

Martin, F., Camfield, L., Rodham, K., Kliempt, P., & Ruta, D. (2007). Twelve years–experience with the Patient Generated Index (PGI) of quality of life: A graded structured review. *Quality of Life Research*, *16*(4), 705–715. https://doi.org/10.1007/s11136-006-9152-6

Mayo, N. E. (2015). *ISOQOL Dictionary of quality of life and health outcomes measurement*. ISOQOL. https://books.google.ca/books?id=cKjksgEACAAJ

Mayo, N. E., Moriello, C., Asano, M., van der Spuy, S., & Finch, L. (2011). The extent to which common health-related quality of life indices capture constructs beyond symptoms and function. *Quality of Life Research*, *20*(5), 621–627. https://doi.org/10.1007/s11136-010-9801-7

McGuire, D., Garrison, L., Armon, C., Barohn, R., Bryan, W., Miller, R., Parry, G., Petajan, J., Ross, M., & The SSNJV/CNTF ALS Study Group. (1996). Relationship of the Tufts Quantitative Neuromuscular Exam (TQNE) and the Sickness Impact Profile (SIP) in

measuring progression of ALS. *Neurology*, *46*(5), 1442–1442. https://doi.org/10.1212/WNL.46.5.1442

McGuire, D., Garrison, L., Armon, C., Barohn, R. J., Bryan, W. W., Miller, R., Parry, G., Petajan, J. H., & Ross, M. A. (1997). A brief quality-of-life measure for ALS clinical trials based on a subset of items from the sickness impact profile. *Journal of the Neurological Sciences*, *152*, s18–s22. https://doi.org/10.1016/S0022-510X(97)00239-6

McHorney, C. A., & Tarlov, A. R. (1995). Individual-patient monitoring in clinical practice: Are available health status surveys adequate? *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *4*(4), 293–307.

Meeberg, G. A. (1993). Quality of life: A concept analysis. *Journal of Advanced Nursing*, *18*(1), 32–38. https://doi.org/10.1046/j.1365-2648.1993.18010032.x

Menzel, P., Dolan, P., Richardson, J., & Olsen, J. A. (2002). The role of adaptation to disability and disease in health state valuation: A preliminary normative analysis. *Social Science & Medicine*, *55*(12), 2149–2158. https://doi.org/10.1016/S0277-9536(01)00358-6

Misajon, R., Hawthorne, G., Richardson, J., Barton, J., Peacock, S., Iezzi, A., & Keeffe, J. (2005). Vision and Quality of Life: The Development of a Utility Measure. *Investigative Opthalmology & Visual Science*, *46*(11), 4007. https://doi.org/10.1167/iovs.04-1389

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, *63*(7), 737–745. https://doi.org/10.1016/j.jclinepi.2010.02.006

Montes, J., Levy, G., Albert, S., Kaufmann, P., Buchsbaum, R., Gordon, P. H., & Mitsumoto, H. (2006). Development and evaluation of self-administered version of the ALSFRS-R. *Neurology*, *67*, 1294–1296. https://doi.org/10.1212/01.wnl.0000238505.22066.fc

Neudert, C., Wasner, M., & Borasio, G. D. (2004). Individual Quality of Life is not Correlated with Health-Related Quality of Life or Physical Function in Patients with Amyotrophic Lateral Sclerosis. *Journal of Palliative Medicine*, *7*(4), 551–557. https://doi.org/10.1089/jpm.2004.7.551

Neudert C., Wasner M., & Borasio G.D. (2001). Patients' assessment of quality of life instruments: A randomised study of SIP, SF-36 and SEIQoL-DW in patients with amyotrophic lateral sclerosis. *Journal of the Neurological Sciences*, *191*(1–2), 103–109. https://doi.org/10.1016/S0022-510X%2801%2900612-8

O'Boyle, C. (1994). The Schedule for the Evaluation of Individual Quality of Life (SEIQoL). Administration Manual. *International Journal of Mental Health*, *23*, 3–23. https://doi.org/10.1080/00207411.1994.11449285

O'Boyle, C., McGee, H., Hickey, A., O'Malley, K., & Joyce, C. R. B. (1992). Individual quality of life in patients undergoing hip replacement. *The Lancet*, *339*(8801), 1088–1091. https://doi.org/10.1016/0140-6736(92)90673-Q

O'Brien, B. J., Spath, M., Blackhouse, G., Severens, J. L., Dorian, P., & Brazier, J. (2003). A view from the bridge: Agreement between the SF-6D utility algorithm and the Health Utilities Index. *Health Economics*, *12*(11), 975–981. https://doi.org/10.1002/hec.789

Ojelabi, A. O., Graham, Y., Haighton, C., & Ling, J. (2017). A systematic review of the application of Wilson and Cleary health-related quality of life model in chronic diseases.

Health and Quality of Life Outcomes, *15*(1), 241. https://doi.org/10.1186/s12955-017-0818-2

Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology*, *46*(1), 1–18. https://doi.org/10.1348/014466506X96931

Pasinelli, P., & Brown, R. H. (2006). Molecular biology of amyotrophic lateral sclerosis: Insights from genetics. *Nature Reviews Neuroscience*, *7*(9), 710–723. https://doi.org/10.1038/nrn1971

Patrick, D. L., Burke, L. B., Powers, J. H., Scott, J. A., Rock, E. P., Dawisha, S., O'Neill, R., & Kennedy, D. L. (2007). Patient-Reported Outcomes to Support Medical Product Labeling Claims: FDA Perspective. *Value in Health*, *10*, S125–S137. https://doi.org/10.1111/j.1524-4733.2007.00275.x

Patrick, D. L., & Erickson, P. (1993). Assessing health-related quality of life for clinical decision-making. In S. R. Walker & R. M. Rosser (Eds.), *Quality of Life Assessment: Key Issues in the 1990s* (pp. 11–63). Springer Netherlands. https://doi.org/10.1007/978-94-011-2988-6_2

Peasgood, T., Brazier, J., Mukuria, C., & Rowen, D. (2014). *A conceptual comparison of well-being measures used in the UK.* (Policy Report No. 26; Policy Researc Unit in Economic Evaluation of Health and Care Interventions (EEPRU)). University of Sheffield. http://www.eepru.org.uk/wp-content/uploads/2017/11/eepru-report-a-conceptual-comparison-of-well-being-measures-sept-2014-026.pdf

Phukan, J., Elamin, M., Bede, P., Jordan, N., Gallagher, L., Byrne, S., Lynch, C., Pender, N., & Hardiman, O. (2012). The syndrome of cognitive impairment in amyotrophic lateral

sclerosis: A population-based study. *Journal of Neurology, Neurosurgery & Psychiatry*, *83*(1), 102–108. https://doi.org/10.1136/jnnp-2011-300188

Pohar, S. L., & Allyson Jones, C. (2009). The burden of Parkinson disease (PD) and concomitant comorbidities. *Archives of Gerontology and Geriatrics*, *49*(2), 317–321. https://doi.org/10.1016/j.archger.2008.11.006

Prell, T., Gaur, N., Stubendorff, B., Rödiger, A., Witte, O. W., & Grosskreutz, J. (2019). Disease progression impacts health-related quality of life in amyotrophic lateral sclerosis. *Journal of the Neurological Sciences*, *397*, 92–95. https://doi.org/10.1016/j.jns.2018.12.035

Reenen, M. van, & Janssen, B. (2019). *EQ-5D-5L User Guide: Basic information on how to use the EQ-5D-5L instrument*. EuroQol Research Foundation. https://euroqol.org/publications/user-guides

Richardson, J., Atherton Day, N., Peacock, S., & Iezzi, A. (2004). Measurement of the Quality of Life for Economic Evaluation and the Assessment of Quality of Life (AQoL) Mark 2 Instrument. *The Australian Economic Review*, *37*(1), 62–88. https://doi.org/10.1111/j.1467-8462.2004.00308.x

Richardson, J., Iezzi, A., Khan, M. A., & Maxwell, A. (2014). Validity and Reliability of the Assessment of Quality of Life (AQoL)-8D Multi-Attribute Utility Instrument. *The Patient - Patient-Centered Outcomes Research*, *7*(1), 85–96. https://doi.org/10.1007/s40271-013-0036-x

Richardson, J., Monash University, Centre for Health Economics, Monash University, & Faculty of Business and Economics. (2011). *Modelling the utility of health states with the assessment of quality of life (AQoL) 8D instrument: Overview and utility scoring algorithm*.

Monash University, Business and Economics, Centre for Health Economics. http://www.buseco.monash.edu.au/centres/che/pubs/researchpaper63.pdf

Richardson, J., Peacock, S. J., Hawthorne, G., Iezzi, A., Elsworth, G., & Day, N. A. (2012). Construction of the descriptive system for the Assessment of Quality of Life AQoL-6D utility instrument. *Health and Quality of Life Outcomes*, *10*, 38. https://doi.org/10.1186/1477-7525-10-38

Robbins, R. A., Simmons, Z., Bremer, B. A., Walsh, S. M., & Fischer, S. (2001). Quality of life in ALS is maintained as physical function declines. *Neurology*, *56*(4), 442–444. https://doi.org/10.1212/WNL.56.4.442

Rooney, J., Burke, T., Vajda, A., Heverin, M., & Hardiman, O. (2017). What does the ALSFRS-R really measure? A longitudinal and survival analysis of functional dimension subscores in amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, *88*(5), 381–385. https://doi.org/10.1136/jnnp-2016-314661

Rosenbaum, P. L., Livingston, M. H., Palisano, R. J., Galuppi, B. E., & Russell, D. J. (2007). Quality of life and health-related quality of life of adolescents with cerebral palsy. *Developmental Medicine & Child Neurology*, *49*(7), 516–521. https://doi.org/10.1111/j.1469-8749.2007.00516.x

Ruta, D. A., & Garratt, A. M. (2013). Health status to quality of life measurement. In C. Jenkinson (Ed.), *Measuring health and medical outcomes* (pp. 138–159). Routledge. https://www.taylorfrancis.com/books/e/9781315072159

Ruta, D. A., Garratt, A. M., Leng, M., Russell, I. T., & MacDonald, L. M. (1994). A New Approach to the Measurement of Quality of Life: The Patient-Generated Index. *Medical Care*, *32*(11), 1109–1126. https://doi.org/10.1097/00005650-199411000-00004

Saarni, S. I., Härkänen, T., Sintonen, H., Suvisaari, J., Koskinen, S., Aromaa, A., & Lönnqvist, J. (2006). The Impact of 29 Chronic Conditions on Health-related Quality of Life: A General Population Survey in Finland Using 15D and EQ-5D. *Quality of Life Research*, *15*(8), 1403–1414. https://doi.org/10.1007/s11136-006-0020-1

Salomon, J. A. (2014). Valuing Health States, Techniques for. In *Encyclopedia of Health Economics* (pp. 454–458). Elsevier. https://doi.org/10.1016/B978-0-12-375678-7.00502-2

Schrooten, M., Smetcoren, C., Robberecht, W., & Van Damme, P. (2011). Benefit of the Awaji diagnostic algorithm for amyotrophic lateral sclerosis: A prospective study. *Annals of Neurology*, *70*(1), 79–83. https://doi.org/10.1002/ana.22380

Schultz, J. (2018). Disease-modifying treatment of amyotrophic lateral sclerosis. *The American Journal of Managed Care*, *24*(15 Suppl), S327–S335.

Seiber, W. J., Groessl, E. J., David, K. M., Ganiats, T. G., & Kaplan, R. M. (2008). *Quality of Well-Being Self-Administered Scale (QWB-SA)*. San Diego Health Services Research Centre, University of California. https://www.researchgate.net/profile/Kristin_Kistler/publication/252316672_Quality_of_well_being_self-administered_QWB-SA_scale/links/5437d6990cf2590375c55a65/Quality-of-well-being-self-administered-QWB-SA-scale.pdf

Silani, V., Messina, S., Poletti, B., Morelli, C., Doretti, A., Ticozzi, N., & Maderna, L. (2011). The diagnosis of Amyotrophic lateral sclerosis in 2010. *Archives Italiennes De Biologie*, *149*(1), 5–27. https://doi.org/10.4449/aib.v149i1.1260

Simmons, Z. (2015). Patient-Perceived Outcomes and Quality of Life in ALS. *Neurotherapeutics*, *12*(2), 394–402. https://doi.org/10.1007/s13311-014-0322-x

Simmons, Z., Bremer, B. A., Robbins, R. A., Walsh, S. M., & Fischer, S. (2000). Quality of life in ALS depends on factors other than strength and physical function. *Neurology*, *55*(3), 388–392. https://doi.org/10.1212/wnl.55.3.388

Simmons, Z., Felgoise, S. H., Bremer, B. A., Walsh, S. M., Hufford, D. J., Bromberg, M. B., David, W., Forshew, D. A., Heiman-Patterson, T. D., Lai, E. C., & McCluskey, L. (2006). The ALSSQOL: Balancing physical and nonphysical factors in assessing quality of life in ALS. *Neurology*, *67*(9), 1659–1664. https://doi.org/10.1212/01.wnl.0000242887.79115.19

Simon Pickard, A., Johnson, J. A., & Feeny, D. H. (2005). Responsiveness of generic health-related quality of life measures in stroke. *Quality of Life Research*, *14*(1), 207–219. https://doi.org/10.1007/s11136-004-3928-3

Sintonen, H. (1981). An approach to measuring and valuing health states. *Social Science & Medicine. Part C: Medical Economics*, *15*(2), 55–65. https://doi.org/10.1016/0160-7995(81)90019-8

Sintonen, H. (1994). *The 15-D measure of health-related quality of life: Reliability, validity and sensitivity of its health state descriptive system*. National Centre for Health Program Evaluation.

Sintonen, H. (1995). *The 15D-Measure of Health-Related Quality of Life. II. Feasibility, Reliability and Validity of Its Valuation System*. National Centre for Health Program Evaluation.

Sintonen, H. (2001). The 15D instrument of health-related quality of life: Properties and applications. *Annals of Medicine*, *33*(5), 328–336. https://doi.org/10.3109/07853890109002086

Sintonen, H., & Pekurinen, M. (1989). *A generic 15 dimensional measure of health-related quality of life (15D)*. *26*, 85–96.

Smith, P. S., Crossley, B., Greenberg, J., Wilder, C., & Carroll, B. (2000). Agreement among three quality of life measures in patients with ALS. *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders : Official Publication of the World Federation of Neurology, Research Group on Motor Neuron Diseases*, *1*(4), 269–275.

Stavem, K., Bjørnæs, H., & Lossius, M. I. (2001). Properties of the 15D and EQ-5D utility measures in a community sample of people with epilepsy. *Epilepsy Research*, *44*(2–3), 179–189. https://doi.org/10.1016/S0920-1211(01)00201-7

Stewart, A. L., & Ware, J. E. (Eds.). (1992). *Measuring functioning and well-being: The medical outcomes study approach*. Duke University Press.

Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use* (Fifth edition). Oxford University Press.

Torrance, G. W. (1987). Utility approach to measuring health-related quality of life. *Journal of Chronic Diseases*, *40*(6), 593–600. https://doi.org/10.1016/0021-9681(87)90019-1

Torrance, G. W., Boyle, M. H., & Horwood, S. P. (1982). Application of Multi-Attribute Utility Theory to Measure Social Preferences for Health States. *Operations Research*, *30*(6), 1043–1069. https://doi.org/10.1287/opre.30.6.1043

Torrance, G. W., Feeny, D. H., Furlong, W. J., Barr, R. D., Zhang, Y., & Wang, Q. (1996). Multiattribute Utility Function for a Comprehensive Health Status Classification System: Health Utilities Index Mark 2. *Medical Care*, *34*(7), 702–722. https://doi.org/10.1097/00005650-199607000-00004

Traynor, B. J., Codd, M. B., Corr, B., Forde, C., Frost, E., & Hardiman, O. M. (2000). Clinical Features of Amyotrophic Lateral Sclerosis According to the El Escorial and Airlie House Diagnostic Criteria: A Population-Based Study. *Archives of Neurology*, *57*(8), 1171. https://doi.org/10.1001/archneur.57.8.1171

Turner, M. R., & Talbot, K. (2013). Mimics and chameleons in motor neurone disease. *Practical Neurology*, *13*(3), 153–164. https://doi.org/10.1136/practneurol-2013-000557

Ubel, P. A., Loewenstein, G., & Jepson, C. (2003). Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Quality of Life Research*, *12*(6), 599–607. https://doi.org/10.1023/A:1025119931010

Valko, K., & Ciesla, L. (2019). Amyotrophic lateral sclerosis. In *Progress in Medicinal Chemistry* (Vol. 58, pp. 63–117). Elsevier. https://doi.org/10.1016/bs.pmch.2018.12.001

van den Berg, L. H., Sorenson, E., Gronseth, G., Macklin, E. A., Andrews, J., Baloh, R. H., Benatar, M., Berry, J. D., Chio, A., Corcia, P., Genge, A., Gubitz, A. K., Lomen-Hoerth, C., McDermott, C. J., Pioro, E. P., Rosenfeld, J., Silani, V., Turner, M. R., Weber, M., … for the Airlie House ALS Clinical Trials Guidelines Group. (2019). Revised Airlie House consensus guidelines for design and implementation of ALS clinical trials. *Neurology*, *92*(14), e1610–e1623. https://doi.org/10.1212/WNL.0000000000007242

van Groenestijn, A. C., Kruitwagen-van Reenen, E. T., Visser-Meily, J. M. A., van den Berg, L. H., & Schröder, C. D. (2016). Associations between psychological factors and health-related quality of life and global quality of life in patients with ALS: A systematic review. *Health and Quality of Life Outcomes*, *14*(1), 107. https://doi.org/10.1186/s12955-016-0507-6

Walters, S. J., Morrell, C. J., & Dixon, S. (1999). Measuring health-related quality of life in patients with venous leg ulcers. *Quality of Life Research*, *8*(4), 327–336. https://doi.org/10.1023/A:1008992006845

Wang, Q., & Chen, G. (1999). The health status of the Singaporean population as measured by a multi-attribute health status system. *Singapore Medical Journal*, *40*(6), 389–396.

Ware, J. E., Kosinski, M., & Keller, S. D. (1996). A 12-Item Short-Form Health Survey: Construction of Scales and Preliminary Tests of Reliability and Validity. *Medical Care*, *34*(3), 220–233. JSTOR. www.jstor.org/stable/3766749

Ware, J. E., Snow, K. K., Kosinski, M., & Gandek, B. (1993). *SF-36 Health Survey Manual and Interpretation Guide*.

Wettergren, L., Kettis-Lindblad, Å., Sprangers, M., & Ring, L. (2009). The use, feasibility and psychometric properties of an individualised quality-of-life instrument: A systematic review of the SEIQoL-DW. *Quality of Life Research*, *18*(6), 737–746. https://doi.org/10.1007/s11136-009-9490-2

Wijesekera, L. C., & Leigh, P. N. (2009). Amyotrophic lateral sclerosis. *Orphanet Journal of Rare Diseases*, *4*(1), 3. https://doi.org/10.1186/1750-1172-4-3

Wilson, I. B. (1995). Linking Clinical Variables With Health-Related Quality of Life: A Conceptual Model of Patient Outcomes. *JAMA*, *273*(1), 59. https://doi.org/10.1001/jama.1995.03520250075037

World Health Organization (Ed.). (2001). *International classification of functioning, disability and health: ICF*. World Health Organization.

Xie, F., Pullenayegum, E., Gaebel, K., Bansback, N., Bryan, S., Ohinmaa, A., Poissant, L., Johnson, J. A., & Canadian EQ-5D-5L Valuation Study Group. (2016). A Time Trade-off-

derived Value Set of the EQ-5D-5L for Canada. *Medical Care*, *54*(1), 98–105. https://doi.org/10.1097/MLR.0000000000000447

Zarei, S., Carr, K., Reiley, L., Diaz, K., Guerra, O., Altamirano, P. F., Pagani, W., Lodin, D., Orozco, G., & Chinea, A. (2015). A comprehensive review of amyotrophic lateral sclerosis. *Surgical Neurology International*, *6*, 171. https://doi.org/10.4103/2152-7806.169561

Chapter Two: Psychometric Properties of Preference-Based Measures for Economic Evaluation in Amyotrophic Lateral Sclerosis: A Systematic Review

Submitted to: *Quality of Life Research*

Psychometric Properties of Preference-Based Measures for Economic Evaluation in Amyotrophic Lateral Sclerosis: A Systematic Review

Nicole Peters, BSc[1], Vanina Dal Bello-Haas, PT, PhD[1], Tara Packham, OTReg(Ont.), PhD[1], Ava Mehdipour, BSc[1], Ayse Kuspinar, PT, PhD[1]

[1]*School of Rehabilitation Science, McMaster University, Hamilton, Ontario, Canada*

ORCID Numbers:

Nicole Peters: 0000-0001-6112-8436

Vanina Dal Bello-Haas: 0000-0003-4066-156X

Tara Packham: 0000-0002-5593-1975

Ava Mehdipour: 0000-0001-7984-9127

Ayse Kuspinar: 0000-0002-3886-0407

Corresponding author:

Ayse Kuspinar, PT, PhD
Assistant Professor
School of Rehabilitation Science
McMaster University
1400 Main St. W. Room 435, IAHS
Hamilton, ON, L8S 1C7
Telephone: 905-525-9140 ext. 27823
kuspinaa@mcmaster.ca

Psychometric Properties of Preference-Based Measures for Economic Evaluation in Amyotrophic Lateral Sclerosis: A Systematic Review

Abstract

Objective: The aim of this review was to synthesize the psychometric properties of generic preference-based measures (PBM) of health-related quality of life (HRQL) in Amyotrophic Lateral Sclerosis (ALS).

Methods: A systematic review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines. Four databases were searched from inception to April 2019: OVID Medline, Embase, PsycINFO and CINAHL. Studies were included if: 1) the sample represented individuals with ALS, 2) a generic PBM was utilized and reported on, and 3) information on the psychometric property of a generic PBM was provided.

Results: Ninety-one articles were screened, and 39 full-text articles were reviewed. Seven full-text articles were included in this review. The mean age of participants ranged from 58.1 to 63.8 years and mean time since diagnosis ranged from 20.5 to 44.6 months. Two generic PBMs were found, the EQ-5D-3L (n=6) and the Quality of Well Being Self-Administered (QWB-SA) scale (n=1). Convergent validity of the EQ-5D-3L was large against a global scale of self-perceived health (r=0.60) and small to large against ALS specific HRQL measures (r=0.19 to 0.75). For the QWB-SA scale, correlations were small against a generic measure (r=0.21), and large against ALS specific measures (r=0.55). The EQ-5D-3L discriminated across different disease severity, however floor effects were reported.

Conclusion: This review highlights the need for more rigorously designed studies to assess the psychometric properties of generic PBMs in ALS and the development of an ALS specific PBM that adequately reflects the health concerns of individuals with ALS.

Keywords

Amyotrophic Lateral Sclerosis, Cost-Utility Analysis, Health-Related Quality of Life, Patient Reported Outcome Measures, Psychometrics

Introduction

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease characterised by selective and progressive degeneration of voluntary motor neurons [1]. Adults with ALS have an overall mortality rate of 80% within the first 2 to 5 years after diagnosis, and experience wide variability in disease severity and disease progression [2]. The disease affects more than 200,000 people worldwide in mid to late adulthood with an average age of onset of 55-66 years [3]. Signs and symptoms of ALS include: a) muscle weakness and atrophy resulting in loss of muscle control; b) spasticity; c) bulbar symptoms such as speech and swallowing difficulties; and, d) respiratory symptoms [4]. With disease progression and the resulting symptoms and loss of independence, the health-related quality of life (HRQL) of individuals with ALS is severely impacted [4–7].

HRQL instruments provide a structured way of including the patient's perspective when evaluating the influence of a disease and its treatments on one's physical, mental and social well-being [5, 7, 8]. HRQL can be assessed using health profiles or preference-based measures (PBMs; also known as utility measures). Health profiles, such as the ALS Specific Quality of Life-Revised (ALSSQOL-R) scale, are scored by subscales and do not produce a single index score useful for economic evaluation purposes [5, 9, 10]. PBMs, on the other hand, are scored from 0.0 (death) to 1.0 (full health), and provide a single value of HRQL [9]. They can be used by researchers and policymakers for economic decision-making purposes to calculate quality-adjusted life years (QALYs) and determine the cost-effectiveness of interventions in ALS [9].

Existing PBMs used with individuals with ALS are generic and consist of measures such as the Short Form 6 Dimension (SF-6D) [11], Health Utilities Index Mark 3 (HUI3) [12] and EuroQol 5 Dimension (EQ-5D) (3 and 5 levels) [13, 14]. For some conditions, such as

rheumatoid arthritis [15], cardiovascular disease [16] and various cancers [17], these measures have established estimates of reliability and validity. However, the reliability and validity of PBMs have not yet been summarized for ALS. As these measures were not developed specifically for individuals with ALS, it is important to assess their psychometric properties in this population [18]. This will assist in understanding whether the values obtained by the scoring system are valid and can be utilized by researchers and policy makers for clinical and cost-evaluation purposes. Therefore, the aim of this review was to synthesize the psychometric properties of generic PBMs of HRQL in ALS.

Methods

A structured search was conducted in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [19] reporting guidelines to identify possible articles that report information on the psychometric properties of PBMs of HRQL in ALS. COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidelines for systematic reviews of patient-reported outcome measures [20] was used to facilitate the understanding of a systematic review on PBMs and determine the quality of PBMs used.

Search Strategy

A research librarian (McMaster University, Hamilton, ON) was consulted for search strategy assistance a priori. Subsequently, a systematic search was conducted to identify all generic PBMs used in ALS. Four databases were searched: OVID Medline (1946 to April 9, 2019), Embase (1974 to April 9, 2019), PsycINFO (1806 to April 2019) and Cumulative Index to Nursing and Allied Health Literature (CINAHL, 1981 to April 9, 2019). Search terms were

related to: i) Amyotrophic Lateral Sclerosis (ALS) AND ii) a generic PBM: EuroQol Five Dimension (EQ-5D) (3 or 5 level), Health Utilities Index (HUI) (Mark 1, 2 or 3), SF-6D, the Assessment of Quality of Life (AQOL), 15-Dimension (15D) or Quality of Well-being (QWB) scale. Medical subject heading (MeSH) search terms and keywords were used for all databases and modified in accordance with the individual database search stipulations. See Supplementary File - Table 1 for the complete search strategy.

Study Selection

Two independent reviewers (NP and AM) identified potentially relevant articles by systematically screening titles/abstracts and then selecting full-text articles for inclusion. Reasons for exclusion were recorded, and if present, differences in responses between the two reviewers were discussed and a consensus reached. A third reviewer (AK) was consulted if a consensus was not reached. Studies were included if: 1) the study sample represented individuals with ALS, 2) a generic PBM of HRQL was utilized and reported on, and 3) potentially relevant information on the psychometric property of a generic PBM was provided, whether this was their objective or not. Only full text articles written in English or French and published in peer-reviewed journals were included in the review. Grey literature, conference proceedings and abstracts were excluded.

Data Extraction

The following information was extracted independently, by two reviewers (NP and AM), from the full text articles selected for data extraction: i) study characteristics - author(s), year of publication, study design, study purpose and study setting, ii) sample characteristics - sample size (*N*), age, gender, time since diagnosis (months), ALS diagnosis, and disease severity, iii)

PBM(s) used (mean ± standard deviation (SD)), and iv) psychometric properties. Specifically, the following metrics were sought from the included articles:

- Reliability - Test-retest reliability: the extent to which scores of a measure have not changed over time, provided the characteristics being measured do not change [21, 22].
- Content validity: the degree to which the content of an instrument is an adequate reflection of the construct of interest [21].
- Construct validity
    - Convergent validity: the degree to which scores of two measurement instruments relate when measuring a similar construct of interest [21, 23].
    - Discriminative (known-groups) validity: the degree to which an instrument is able to discriminate between two groups that differ on the construct being measured [24].
- Predictive validity: the extent to which measurement instrument scores are an adequate reflection of a gold standard for the construct of interest in the future [18].
- Responsiveness: the ability of an instrument to detect change over time in the construct of interest [21].
- Floor/ceiling effect: the percentage of the sample obtaining scores at the lower and upper ends of the scale, respectively [18]; known as a form of interpretability that can affect the responsiveness of an instrument [18]. Floor and ceiling effects were deemed significant when percentage values >15% were seen [25].

Evaluation of Measurement Properties

The evaluation of measurement properties consisted of three steps. First, the methodological quality of studies was assessed using the relevant boxes for each measurement

property included in the COSMIN Risk of Bias Checklist [26]. Second, the results of each study were rated against COSMIN's criteria for good measurement properties as either sufficient (+), insufficient (-) or indeterminate (?) [26]. Third, all results were rated and graded using COSMIN's modified Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach (Supplementary File - Table 2 & 3) [20, 26]. The evaluation of measurement properties could only be assessed for studies whose primary or secondary objective(s) was to evaluate the psychometric properties of a PBM [27].

The hypotheses derived were used to evaluate the psychometric properties when evaluating results against COSMIN's criteria for good measurement properties [26]. Reliability correlation coefficients were hypothesized to be greater or equal to 0.70 [18]. For measures assessing similar constructs (e.g. HRQL), we hypothesized large correlations of ≥ 0.50 [18, 28]. For measures assessing related, but dissimilar constructs (e.g. function/disease severity), we hypothesized a medium correlation of 0.30-0.49 [18, 28]. For discriminative (known-groups) validity, we hypothesized a significant difference in mean scores ($p < 0.05$) between groups of different pre-determined variables (e.g. ALS severity levels) [26]. For predictive validity, areas under the curve (AUCs) were hypothesized to be greater than or equal to 0.70 [26]. Responsiveness was hypothesized to be significant at $p < 0.05$ or with an AUC ≥ 0.70 [26].

Results

Results of Search

A total of 135 records were identified through the database searches. Forty-four records were removed due to duplication, resulting in a total of 91 articles for screening. Fifty-two articles were excluded during the initial screening of titles and abstracts. From this, 39 full-text

articles were assessed for eligibility, whereby 32 of those articles were subsequently excluded. Articles were excluded if: i) a generic PBM was not assessed (n=4), ii) the psychometric properties of a generic PBM was not assessed (n=8), iii) the study did not report on or assess the population of interest (n=4), and iv) articles were grey literature, conference proceedings or abstracts (n=16). This left seven full-text articles for inclusion in the review. Figure 1 outlines the complete review process.

Sample Characteristics

Table 1 presents key characteristics and psychometric properties from each study included in the review. Sample sizes across the seven studies ranged from 19 to 214 participants and 31% to 49% female. The mean participant age ranged from 58.1 to 63.8 years, and a mean time since ALS diagnosis of 20.5 to 44.6 months. ALS severity was classified according to: i) the ALS Functional Rating Scale-Revised (ALSFRS-R) (mean score = 32.63) [29]; ii) the ALS Severity Scale (ALSSS) (mean score = 27.1) [30]; iii) High or low severity classified as requiring caregiver assistance or not (75% of sample classified as high) [31]; or iv) the ALS Health State Scale (ALS/HSS) (27-29% of sample classified as moderate or severe ALS) [32, 33]. If ALS severity was not reported, ALS diagnosis was classified using the El Escorial criteria with 21% to 47% of the sample classified as probable or definite ALS [34, 35].

Generic Preference-Based Measure(s) Used

Two PBMs were examined in the included studies: the EQ-5D-3L (n=6) [29–33, 35] and the QWB Self-Administered (QWB-SA) scale (n=1) [34]. The EQ-5D-3L is a widely used generic PBM of HRQL [36]. It consists of five domains (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) and produces a single index score for health utility ranging from -0.594 for the worst possible health state to 1.0 for the best possible health state [37]. The

QWB scale is an interview administered scale that has been developed for self-administration (QWB-SA). It combines three scales of functioning with a measure of symptoms and problems and produces a single index score that ranges from 0.0 (death) to 1.0 (full function) [38].

Only one [34] of the seven included studies' primary purpose was to evaluate the psychometric property of a generic PBM, the QWB-SA scale. The remaining six [29–33, 35] studies reported information on the psychometric properties of a generic PBM, the EQ-5D-3L; however, it was not the purpose of their study. Mean EQ-5D-3L scores ranged from 0.18 to 0.54 and a range of 37 to 214 individuals with ALS were included in these studies. A mean QWB-SA score of 0.43 was reported and nineteen individuals with ALS were included in this study [34].

Psychometric Properties

Convergent validity, discriminative (known-groups) validity and floor effects were reported in the seven included studies.

Convergent Validity

For the EQ-5D-3L, convergent validity was evaluated in four out of six studies (Table 2) [29, 30, 33, 35]. A large correlation of 0.60 with the EQVAS was reported in a single study (n=77) [33]. Correlations with a disease-specific health profile, the ALS Assessment Questionnaire 40 (ALSAQ-40) subscales ranged from small with the Eating and Drinking (ALSED) subscale (r=0.19) to large with the Activities of Daily Living and Independence (ALSADL-I) subscale (r=-0.75) [33]. A large correlation of 0.72 with the disease-specific functional measure, ALSFRS-R, was reported in a smaller study (n=46) [29]. A medium correlation of 0.43 was found with social support, as measured by the FSozU K-14 measure [29].

For the QWB-SA scale, one very small study (n=19) evaluated convergent validity against a generic (SF-36) (r=0.21) and disease-specific (Sickness Impact Profile ALS-19 (SIP/ALS-19)) (r=0.55) health profile (Table 2) [34].

Discriminative (Known-Groups) Validity

For the EQ-5D-3L, all six studies evaluated known-groups validity (Table 1). This property was not assessed for the QWB-SA scale. The EQ-5D-3L was able to discriminate between patients across disease severity, with evidence of statistical differences in mean scores [29–33, 35]. Of the three studies including mean values, the mean scores decreased (range = 0.65 to -0.01) with increasing disease severity [31, 33, 35]. Discriminative ability of the EQ-5D-3L was evident between people with bulbar or limb-onset ALS, with bulbar patients reporting a significantly higher EQ-5D-3L score (median = 46.4) than limb-onset patients (median = 14.9) [29]. Known-groups validity was also established against two other neuromuscular diseases (i.e. myasthenia gravis (MG) and facioscapulohumeral muscular dystrophy (FSHD)), with lower scores reported in individuals with ALS [30] compared to individuals with MG and FSHD.

Floor Effects

Floor effects were reported for the EQ-5D-3L, where 54% to 92% of individuals with ALS reported moderate or severe problems across all five dimensions of the measure (Table 1) [29, 30, 32].

Evaluation of Psychometric Properties

Six out of seven studies could not be evaluated on the psychometric properties reported, as only one [34] of the seven studies' primary purpose was to evaluate the psychometric property of a generic PBM. For this study [34], a methodological quality analysis of the data resulted in a serious risk of bias; determined using COSMIN's risk of bias checklist [27]. In grading the

quality of evidence using the GRADE approach and in accordance with hypotheses, there was serious inconsistency, very serious imprecision and serious indirectness. This resulted in an overall rating of very low (Table 1).

Discussion

To our knowledge, this was the first study systematically reviewing the psychometric properties of generic PBMs in ALS. Across the seven studies included in this review, only the EQ-5D-3L and the QWB-SA scale were used in ALS. Furthermore, convergent validity, known-groups validity and floor/ceiling effects were the only psychometric properties assessed for these measures in this population. Our review revealed that other important psychometric properties of PBMs (i.e. content validity, reliability and responsiveness) have not yet been evaluated in ALS. Furthermore, none of the included studies, with one exception, were specifically designed to assess the psychometric properties of a generic PBM in the ALS population [34]. When the methodological quality of this study was assessed, the quality was graded as very low, preventing an accurate conclusion regarding the usability of the QWB-SA scale in the ALS population.

The EQ-5D-3L was highly correlated with the ALSFRS-R, an ALS specific functional rating scale reflective of disease severity; well exceeding our hypothesized correlation of less than 0.5 (for comparison of dissimilar constructs HRQL and disease severity). This is not entirely unexpected however as both the EQ-5D-3L [9] and the ALSFRS-R [39] contain similar domains, such as mobility and self-care, that are highly affected in ALS: this may explain the large correlations observed between the two measures [5]. Moreover, mobility is a domain that is

greatly affected in various conditions, including ALS [40], due to its relation to independence and quality of life. As such, it is often included as a construct in many generic PBMs of HRQL.

The QWB-SA scale, however, may not be a generic measure that can be used in this population due to our study's findings and the unique nature of symptoms experienced by individuals with ALS. For example, the QWB-SA scale contains items that address mobility; however the items are very symptom and limitations focused with little emphasis on ALS-relevant items such as functional mobility, speech or pain [34, 41]. This could result in items that are not relevant to this population or even an underrepresentation of items that are. Furthermore, the structure of the QWB-SA scale includes a style of item weighting that results in items relevant to individuals with ALS to contribute much less to the overall score. Furthermore, the QWB-SA scale was shown to weakly correlate with the generic SF-36 ($r=0.21$) and strongly correlate with the disease specific SIP/ALS-19 health profile ($r=0.55$). Respectively, a correlation ≥ 0.50 and a correlation of 0.30-0.49 would be expected, however the opposite was observed. Additionally, this was the only study included in the review with the primary purpose of psychometric evaluation. When the quality of evidence was assessed, it was deemed to be poor [27, 34]. As the QWB-SA scale was observed to correlate weakly with certain domains of the SF-36 that were similar in the EQ-5D-3L and the ALSFRS-R, items included in the QWB-SA scale may not truly capture what is important to individuals with ALS or be the best tool for use in this population. However, as only one study has assessed this, further research is warranted in order to be make accurate recommendations.

At a total score level, the EQ-5D-3L measure in ALS was able to discriminate between patients across disease severity as evidenced by significant differences in mean scores. However, at the individual item level, there is a prominent floor effect as majority of individuals reported

moderate or severe problems in EQ-5D-3L domains; indicating the full scope of the disease is not being captured. This can affect the responsiveness of an instrument and the ability to accurately detect change over time [18]. For individuals with ALS, this is important to take note of as responsiveness is a critical property for assessing the cost-effectiveness of interventions in ALS [9]. Moreover, content validation, a fundamental component of validity, was not assessed in any of the studies. As such, generic PBMs may miss domains that are important or specific to individuals with ALS. For example, valued domains such as recreation and leisure activities, and interpersonal relationships have been identified by individuals with ALS to be important to their quality of life [8]. However, these domains are not always assessed by generic PBMs. The development of an ALS specific PBM would be one possible solution to help ensure that included domains reflect the health concerns of individuals with ALS.

PBMs, such as the EQ-5D-3L and the QWB-SA scale, were developed to provide evidence on the benefits or harms of a treatment on HRQL from the patient's perspective [9]. They provide a single index value of HRQL used to produce QALYs in order to evaluate the cost-effectiveness of interventions for a health condition [9]. PBMs can be of great use to patients, clinicians and researchers alike, however, our results indicate there is limited evidence of their psychometric properties in ALS.

One limitation for this systematic review is the small sample of studies included. As only one study's primary purpose was the psychometric evaluation of a generic PBM, there is limited evidence regarding the psychometric properties of generic PBMs in ALS. Another limitation is the use of only two generic PBMs in ALS; this may result in an imprecise representation and accuracy of generic PBMs' use in ALS.

Conclusion

To our knowledge, this is the first study systematically reviewing the psychometric properties of generic PBMs in ALS. The EQ-5D-3L was the most reported generic PBM. Although this measure demonstrated convergent and known-groups validity in ALS, significant floor effects were observed for all items, indicating that questions may not be appropriate for individuals with ALS. The only other measure used was the QWB-SA scale, which showed poor quality in its assessment of convergent validity and revealed items that are not relevant to individuals with ALS. Furthermore, there were psychometric properties of generic PBMs that have not been assessed in ALS, namely content validity, reliability and responsiveness. Therefore, our results highlight the need for more rigorously designed studies assessing the psychometric properties of generic PBMs in ALS or the development of an ALS specific PBM that reflects the health concerns of individuals with ALS.

Declarations

Funding. This work was supported by the 2018 ALS Project Grant awarded by the ALS Society of Canada.

Authors' contributions. All authors (NP, VDH, TP, AM, AK) contributed to the study conception and design. Screening and data extraction of articles was conducted by NP and AM. The first draft of the manuscript was written by NP. Preliminary edits were made by AK and all authors commented on previous iterations of the manuscript. All authors read and approved the final manuscript and NP is the guarantor of the paper.

Data availability. The data that supports the findings of this study are available in this published article and in the supplementary material of this article.

Compliance with ethical standards

Conflicts of interest. The authors declare that they have no conflicts of interest.

References

1. Rowland, L. P., & Shneider, N. A. (2001). Amyotrophic Lateral Sclerosis. *New England Journal of Medicine*, *344*(22), 1688–1700. https://doi.org/10.1056/NEJM200105313442207
2. Valko, K., & Ciesla, L. (2019). Amyotrophic lateral sclerosis. In *Progress in Medicinal Chemistry* (Vol. 58, pp. 63–117). Elsevier. https://doi.org/10.1016/bs.pmch.2018.12.001
3. Longinetti, E., & Fang, F. (2019). Epidemiology of amyotrophic lateral sclerosis: an update of recent literature. *Current Opinion in Neurology*, *32*(5), 771–776. https://doi.org/10.1097/WCO.0000000000000730
4. Zarei, S., Carr, K., Reiley, L., Diaz, K., Guerra, O., Altamirano, P. F., … Chinea, A. (2015). A comprehensive review of amyotrophic lateral sclerosis. *Surgical Neurology International*, *6*, 171. https://doi.org/10.4103/2152-7806.169561
5. Simmons, Z. (2015). Patient-Perceived Outcomes and Quality of Life in ALS. *Neurotherapeutics*, *12*(2), 394–402. https://doi.org/10.1007/s13311-014-0322-x
6. Handy, C. R., Krudy, C., Boulis, N., & Federici, T. (2011). Pain in Amyotrophic Lateral Sclerosis: A Neglected Aspect of Disease. *Neurology Research International*, *2011*, 1–8. https://doi.org/10.1155/2011/403808
7. Swash, M. (1997). Health outcome and quality-of-life measurements in amyotrophic lateral sclerosis. *Journal of Neurology*, *244*(S2), S26–S29. https://doi.org/10.1007/BF03160578
8. Karimi, M., & Brazier, J. (2016). Health, Health-Related Quality of Life, and Quality of Life: What is the Difference? *PharmacoEconomics*, *34*(7), 645–649. https://doi.org/10.1007/s40273-016-0389-9
9. Brazier, J., Ratcliffe, J., Salomon, J. A., & Tsuchiya, A. (2017). *Measuring and valuing health benefits for economic evaluation* (Second edition.). Oxford: Oxford University Press.

10. McGuire, D., Garrison, L., Armon, C., Barohn, R. J., Bryan, W. W., Miller, R., … Ross, M. A. (1997). A brief quality-of-life measure for ALS clinical trials based on a subset of items from the sickness impact profile. *Journal of the Neurological Sciences*, *152*, s18–s22. https://doi.org/10.1016/S0022-510X(97)00239-6

11. Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*, *21*(2), 271–292. https://doi.org/10.1016/S0167-6296(01)00130-8

12. Horsman, J., Furlong, W., Feeny, D., & Torrance, G. (2003). The Health Utilities Index (HUI): concepts, measurement properties and applications. *Health and Quality of Life Outcomes*, *1*, 54. https://doi.org/10.1186/1477-7525-1-54

13. Brooks, R. (1996). EuroQol: the current state of play. *Health Policy*, *37*(1), 53–72. https://doi.org/10.1016/0168-8510(96)00822-6

14. Herdman, M., Gudex, C., Lloyd, A., Janssen, Mf., Kind, P., Parkin, D., … Badia, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*, *20*(10), 1727–1736. https://doi.org/10.1007/s11136-011-9903-x

15. Bombardier, C., & Raboud, J. (1991). A comparison of health-related quality-of-life measures for rheumatoid arthritis research. *Controlled Clinical Trials*, *12*(4), S243–S256. https://doi.org/10.1016/S0197-2456(05)80028-5

16. Dyer, M. T., Goldsmith, K. A., Sharples, L. S., & Buxton, M. J. (2010). A review of health utilities using the EQ-5D in studies of cardiovascular disease. *Health and Quality of Life Outcomes*, *8*(1), 13. https://doi.org/10.1186/1477-7525-8-13

17. Longworth, L., Yang, Y., Young, T., Mulhern, B., Hernández Alava, M., Mukuria, C., … Brazier, J. (2014). Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. *Health Technology Assessment*, *18*(9). https://doi.org/10.3310/hta18090

18. de Vet, H. C. W. (Ed.). (2011). *Measurement in medicine: a practical guide*. Cambridge ; New York: Cambridge University Press.

19. Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery*, *8*(5), 336–341. https://doi.org/10.1016/j.ijsu.2010.02.007

20. Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, *27*(5), 1147–1157. https://doi.org/10.1007/s11136-018-1798-3

21. Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., … de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, *63*(7), 737–745. https://doi.org/10.1016/j.jclinepi.2010.02.006

22. Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: a practical guide to their development and use* (Fifth edition.). Oxford: Oxford University Press.

23. Chin, C.-L., & Yao, G. (2014). Convergent Validity. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 1275–1276). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_573

24. Davidson, M. (2014). Known-Groups Validity. In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 3481–3482). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-0753-5_1581

25. McHorney, C. A., & Tarlov, A. R. (1995). Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *4*(4), 293–307.

26. Mokkink, L. B., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., De Vet, H. C. W., & Terwee, C. B. (2018). COSMIN methodology for systematic reviews of Patient - Reported Outcome Measures (PROMs). User Manual, (February), 1–78.

27. Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research*, *27*(5), 1171–1179. https://doi.org/10.1007/s11136-017-1765-4

28. Cohen, J. (1988). Statistical Power Analysis for the Behavioural Science (2nd Edition). In *Statistical Power Anaylsis for the Behavioural Science (2nd Edition)*.

29. Ilse, B., Prell, T., Walther, M., Hartung, V., Penzlin, S., Tietz, F., … Grosskreutz, J. (2015). Relationships between disease severity, social support and health-related quality of life in patients with amyotrophic lateral sclerosis. *Social Indicators Research*, *120*(3), 871–882. https://doi.org/10.1007/s11205-014-0621-y

30. Winter, Y., Schepelmann K., Spottke A.E., Claus D., Grothe C., Schroder R., … Dodel R. (2010). Health-related quality of life in ALS, myasthenia gravis and facioscapulohumeral muscular dystrophy. *Journal of Neurology*, *257*(9), 1473–1481. https://doi.org/10.1007/s00415-010-5549-9

31. Lopez-Bastida, J., Perestelo-Perez L., Monton-Alvarez F., Serrano-Aguilar P., & Alfonso-Sanchez J.L. (2009). Social economic costs and health-related quality of life in patients with amyotrophic lateral sclerosis in Spain. *Amyotrophic Lateral Sclerosis*, *10*(4), 237–243. https://doi.org/10.1080/17482960802430781

32. Kiebert, G., Green C., Murphy C., Mitchell J.D., O'Brien M., Burrell A., & Leigh P.N. (2001). Patients' health-related quality of life and utilities associated with different stages of amyotrophic lateral sclerosis. *Journal of the Neurological Sciences*, *191*(1–2), 87–93. https://doi.org/10.1016/S0022-510X%2801%2900616-5

33. Green, C., Kiebert G., Murphy C., Mitchell J.D., O'Brien M., Burrell A., & Leigh P.N. (2003). Patients' health-related quality-of-life and health state values for motor neurone disease/amyotrophic lateral sclerosis. *Quality of Life Research*, *12*(5), 565–574. https://doi.org/10.1023/A:1025052609818

34. Smith, P. S., Crossley, B., Greenberg, J., Wilder, C., & Carroll, B. (2000). Agreement among three quality of life measures in patients with ALS. *Amyotrophic lateral sclerosis and other motor neuron disorders : official publication of the World Federation of Neurology, Research Group on Motor Neuron Diseases*, *1*(4), 269–75.

35. Jones, A.R., Jivraj N., Balendra R., Murphy C., Kelly J., Thornhill M., … Al-Chalabi A. (2014). Health utility decreases with increasing clinical stage in amyotrophic lateral

sclerosis. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *15*(3–4), 285–291. https://doi.org/10.3109/21678421.2013.872149

36. Brazier, J., Ara, R., Rowen, D., & Chevrou-Severac, H. (2017). A Review of Generic Preference-Based Measures for Use in Cost-Effectiveness Models. *PharmacoEconomics*, *35*(S1), 21–31. https://doi.org/10.1007/s40273-017-0545-x
37. Dolan, P. (1997). Modeling Valuations for EuroQol Health States: *Medical Care*, *35*(11), 1095–1108. https://doi.org/10.1097/00005650-199711000-00002
38. Kaplan, R. M., Sieber, W. J., & Ganiats, T. G. (1997). The quality of well-being scale: Comparison of the interviewer-administered version with a self-administered questionnaire. *Psychology & Health*, *12*(6), 783–791. https://doi.org/10.1080/08870449708406739
39. Cedarbaum, J. M., Stambler, N., Malta, E., Fuller, C., Hilt, D., Thurmond, B., & Nakanishi, A. (1999). The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. *Journal of the Neurological Sciences*, *169*(1–2), 13–21. https://doi.org/10.1016/S0022-510X(99)00210-5
40. Kuspinar, A., & Mayo, N. E. (2013). Do generic utility measures capture what is important to the quality of life of people with multiple sclerosis? *Health and Quality of Life Outcomes*, *11*(1), 71. https://doi.org/10.1186/1477-7525-11-71
41. Kaplan, R. M., & Anderson, J. P. (1988). A general health policy model: update and applications. *Health Services Research*, *23*(2), 203–235.
42. Bansback, N., Tsuchiya, A., Brazier, J., & Anis, A. (2012). Canadian Valuation of EQ-5D Health States: Preliminary Value Set and Considerations for Future Valuation Studies. *PLoS ONE*, *7*(2), e31115. https://doi.org/10.1371/journal.pone.0031115

Supplementary File

Table 1. Search strategy used for literature search

Concept	Search Terms
Population	Amyotrophic Lateral Sclerosis/ Amyotrophic lateral sclerosis.mp. ALS.mp. Lou Gehrig* disease*.mp. Motor Neuron Disease*.mp.
Preference-based measures - *Italics* = CINAHL only	EuroQol.mp. EQ-5D*.mp. *EQ 5D*.mp.* EQ-5D-5L.mp. EQ-5D-3L.mp. EQ5D-5L.mp. EQ5D-3L.mp. EQ5D5L.mp. EQ5D3L.mp. *EQ 5D 5L* *EQ 5D 3L* *EQ5D 5L* *EQ5D 3L* Health utilit* index*.mp. HUI.mp. HUI1.mp. HUI2.mp. HUI3.mp. HUI-1.mp. HUI-2.mp. HUI-3.mp. HUI-I.mp. HUI-II.mp. HUI-III.mp. *HUI 1.mp.* *HUI 2.mp.* *HUI 3.mp.* *HUI I.mp.* *HUI II.mp.* *HUI III.mp.* SF-6D*.mp. SF6D*.mp. Short Form 6D.mp. Short Form Six Dimension.mp. *SF 6D*.mp.* *Short-Form 6D.mp.*

	Short-Form Six-Dimension.mp. AQOL.mp. Assessment of quality of life.mp. Generic utility* measure*.mp. Generic preference based measure*.mp. *Generic preference-based measure*.mp.* Preference based measure*.mp. *Preference-based measure*.mp.* Quality of well-being.mp. *Quality of well being.mp.* QWB.mp. 15D*.mp. 15-D*.mp. *15 D.mp.* 15-Dimension*.mp. *15 Dimension*.mp.*

.mp. – keyword designation for databases, * - open ended word for search (e.g. plural, hyphen etc.) to allow for as many possible results as possible

Table 2. COSMIN's modified GRADE approach for grading the quality of evidence [1]

Quality of Evidence	**Lower if**
High[a]	Risk of bias -1 Serious -2 Very serious -3 Extremely serious Inconsistency -1 Serious -2 Very serious Imprecision -1 total n=50-100 -2 total n<50 Indirectness -1 Serious -2 Very serious
Moderate[b]	
Low[c]	
Very Low[d]	

The starting point is the assumption that the quality of evidence is of high quality. The quality of evidence is subsequently downgraded to moderate, low or very low when there is a risk of bias (study quality), inconsistency (unexplained) in results, imprecision (from sample size) or indirect results. Information on the process is described in detail in the COSMIN user manual [1]. Definitions were adapted from the GRADE approach [2].

[a]Very confident that the true measurement property lies close to that of the estimate of the measurement property
[b]Moderately confident in the estimate of the measurement property; it is likely close to the true measurement property
[c]Confidence in the measurement property estimate is limited: it may be substantially different from the true measurement property
[d]Very little confidence in the measurement property estimate: it is likely to be substantially different from the true measurement property
n - sample size

Table 3. Instructions for downgrading risk of bias, inconsistency and indirectness [1]

Risk of bias	**Downgrading for Risk of Bias**
No	There are multiple studies of at least adequate quality, or there is one study of very good quality available
Serious	There are multiple studies of doubtful quality available, or there is only one study of adequate quality
Very serious	There are multiple studies of inadequate quality, or there is only one study of doubtful quality available
Extremely serious	There is only one study of inadequate quality available
Inconsistency*	**Downgrading for Inconsistency**
Serious	If >50% of results were rated as sufficient according to COSMIN's criteria for good measurement properties
Very serious	If <50% of results were rated as sufficient according to COSMIN's criteria for good measurement properties
Indirectness	**Downgrading for Indirectness**
Serious	If other populations were also examined or none of the comparison measures examined quality of life or HRQL (for convergent validity and responsiveness) in the study.
Very serious	If other populations were also examined and none of the comparison measures examined quality of life or HRQL (for convergent validity and responsiveness) in the study.

*only for inconsistent ratings

References

1. Mokkink, L. B., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN methodology for systematic reviews of Patient - Reported Outcome Measures (PROMs). User Manual, (February), 1–78.
2. Scünemann, H., Brożek, J., Guyatt, G., & Oxman, A. (Eds.). (2013). Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Retrieved from https://gdt.gradepro.org/app/handbook/handbook.html

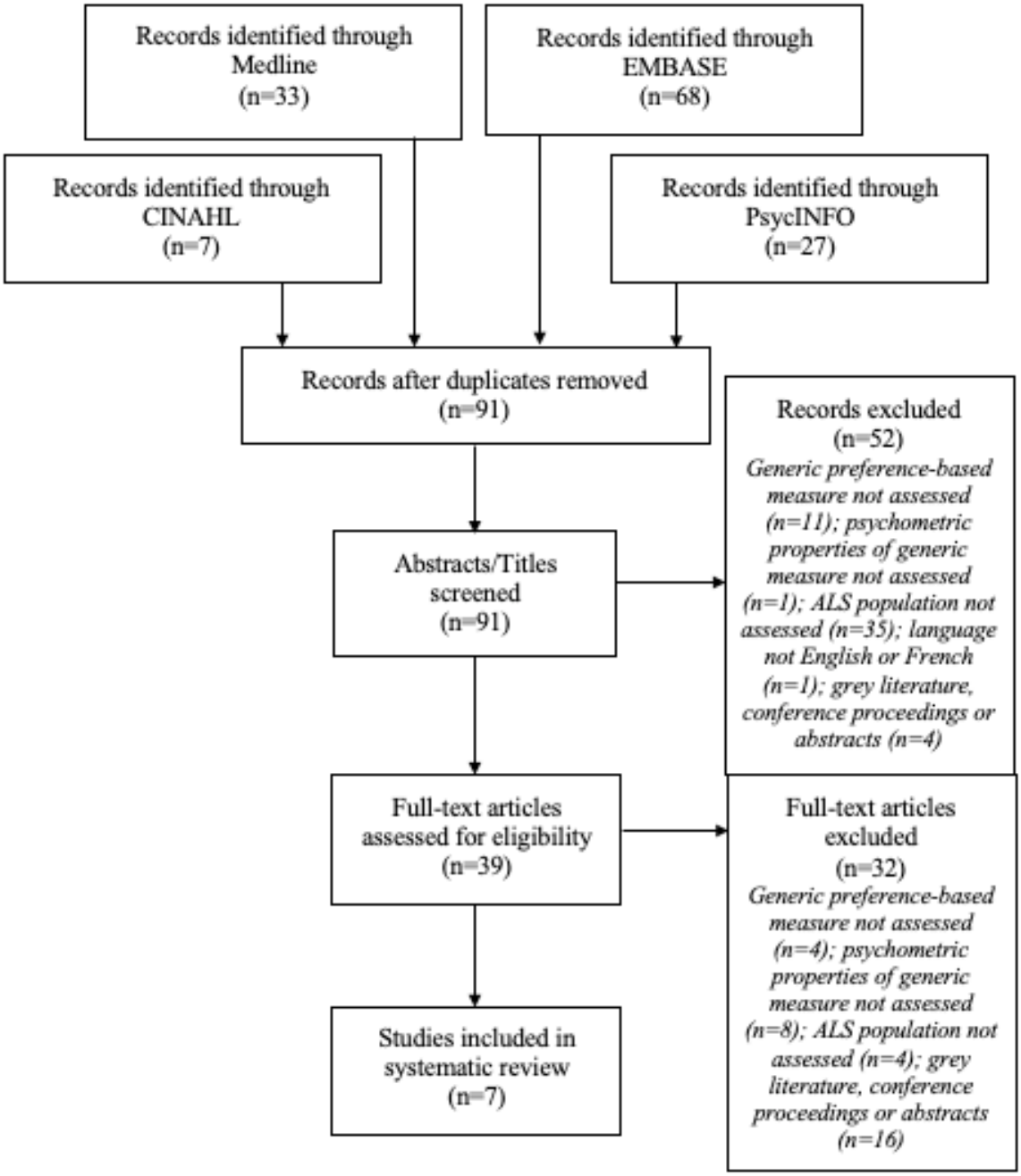


Figure 1. Flow diagram of article selection process. Adapted from the PRISMA statement [18]. *ALS* Amyotrophic Lateral Sclerosis.

Table 1. Description of included studies.

Author (year)	Country	Study design	Study purpose	Study setting	Sample size (*N*)	Sample characteristics	ALS severity or diagnosis	Preference-based measure used	Mean ± SD for measure	Known-groups validity	Convergent validity	Floor effect	Evaluation of measurement properties
Ilse et al. (2015) [29]	Germany	Cross-sectional study	To describe the relationship between HRQL using the EQ-5D, disease severity and social support in patients with ALS	Outpatient clinic	*N* = 49	Age 63.8 ± 10.0, 49% Female, Disease duration 35.1 months ± 36.3, Time since diagnosis not presented, ALSFRS-R 32.6 ± 9.2 (range 0-48)	Severity classified according to ALSFRS-R 32.6 ± 9.2 (range 0-48)	EQ-5D-3L, EQVAS	EQ-5D* 0.36 ± 0.29 EQVAS 42.8 ± 24.1	Bulbar-onset patients had a significantly higher EQ-5D score (median 46.4) than limb-onset patients (median 14.9) (p=0.034)	EQ-5D[a]: +FSozU K-14[b] (r=0.43, p=0.087), +BDI (r=-0.43), +ALSFRS-R (r=0.72, p<0.001)	61-86% of individuals with ALS reported moderate/severe problems in EQ-5D dimensions compared to 3-28% in general population	Not assessed – primary objective was not to assess psychometric properties of generic preference-based measure
Jones et al. (2014) [35]	UK	Longitudinal clinical trial	To assess whether ALS clinical staging could be used in cost-effectiveness analyses	10 outpatient clinics	*N* = 214	Age 58.1 ± 10.8, 31% Female, Time since diagnosis not presented, ALSFRS-R score not presented.	ALS severity not presented, Diagnosis classified according El Escorial criteria: Definite (n=82, 38%), Probable (n=80, 37%), Probable laboratory supported ALS (n=38, 17%), Possible (n=14, 5%)	EQ-5D-3L, EQVAS	EQ-5D score not presented for total sample	Mean EQ-5D scores decreased with increasing ALS severity[c]: from 0.65 (less severe) to 0.27 (more severe) (p<0.001)	ALS clinical stage is a predictor of EQ5D score (χ^2 145.08, $p=3.14x10^{32}$)	Not available	Not assessed – primary objective was not to assess psychometric properties of generic preference-based measure

Author (year)	Country	Study design	Study purpose	Study setting	Sample size (*N*)	Sample characteristics	ALS severity or diagnosis	Preference-based measure used	Mean ± SD for measure	Known-groups validity	Convergent validity	Floor effect	Evaluation of measurement properties
Winter et al. (2010) [30]	Germany	Cross-sectional study	To compare HRQL in patients with ALS, FSHD and MG and to identify the determinants of HRQL in each disease	7 outpatient clinics	Total *N* = 91, ALS *N* = 37	Age 59.6 ± 11.0, 43% Female, Time since diagnosis 39.7 months ± 73.7, ALSSS 27.1 ± 6.8 (range 0-40)	Severity classified according to the ALSSS 27.1 ± 6.8 (range 0-40)	EQ-5D-3L, EQVAS	EQ-5D* 0.54 ± 0.32 (median 0.70), EQVAS 0.38 ± 0.15 (median 0.40)	Mean EQ-5D scores were significantly lower in ALS (0.54) compared to FSHD (0.75) and MG (0.89) (p<0.01)	ALSSS significantly associated with EQ-5D (p <0.01)	70-92% of individuals with ALS reported moderate/severe problems in EQ-5D dimensions compared to 3- 28% in general population	Not assessed – primary objective was not to assess psychometric properties of generic preference-based measure
López-Bastida et al. (2009) [31]	Spain	Cross-sectional study	To determine the economic burden (direct and indirect costs) and assess HRQL in patients with ALS in Spain	Multiple outpatient clinics across 7 regions	*N* = 63	Age 59.1 ± 10.3, 48% Female, Time since diagnosis 44.6 months ± 62.4, ALSFRS-R score not presented.	Severity classified according to High[d] or Low[e] severity: High severity (n=47, 75%), Low severity (n=16, 25%)	EQ-5D-3L, EQVAS	EQ-5D* 0.18 ± 0.22, EQVAS 0.29 ± 0.23	High severity: EQ-5D 0.12 ± 0.17, EQVAS 26 ± 22 Low severity: EQ-5D 0.35 ± 0.27, EQVAS 38 ± 23 EQ-5D scores decreased with increasing ALS severity from 0.35 to 0.12 (p <0.05)	Not available	Not available	Not assessed – primary objective was not to assess psychometric properties of generic preference-based measure
Green et al. (2003) [33]	UK	Cross-sectional study	To examine the relationship between disease severity, HRQL and health state values in patients with MND	2 outpatient clinics	*N* = 77	Age 58.1 ± 12.1 (range 27-79), 36% Female, Time since diagnosis 25.3 months ± 22.6 (range 1-112), ALSFRS-R score not presented.	Severity classified according to the ALS/HSS: Level 1 Mild (n=15, 20%), Level 2 Moderate (n=21, 27%), Level 3 Severe (n=22, 29%), Level 4 Terminal (n=19, 25%)	EQ-5D-3L, EQVAS	EQ-5D* 0.35 ± 0.35, (95% CI 0.27-0.43) (median 0.31), EQVAS 0.55 ± 0.22 (95% CI 0.5-0.6) (median 0.50)	Mean EQ-5D scores decreased with increasing ALS severity from 0.63 to -0.01 (p<0.015)	EQ-5D[f]: +EQVAS (r=0.60[g]); +ALSPM (r=-0.60[g]), +ALSADL/I (r=-0.75[g]), +ALSED (r=0.19[g]), +ALSCOM (r=-0.32[g]), +ALSER (r=-0.43[g])	Not available	Not assessed – primary objective was not to assess psychometric properties of generic preference-based measure

Author (year)	Country	Study design	Study purpose	Study setting	Sample size (*N*)	Sample characteristics	ALS severity or diagnosis	Preference-based measure used	Mean ± SD for measure	Known-groups validity	Convergent validity	Floor effect	Evaluation of measurement properties
Kiebert et al. (2001) [32]	UK	Cross-sectional study	To assess HRQL and health state values in a sample of patients with different levels of severity of ALS	2 outpatient clinics	*N* = 77	Age 58.1 ± 12.1 (range 27-79), 36% Female, Time since diagnosis 25.3 months ± 22.6 (range 1-112), ALSFRS-R score not presented.	Severity classified according to the ALS/HSS: Level 1 Mild (n=15, 20%), Level 2 Moderate (n=21, 27%), Level 3 Severe (n=22, 29%), Level 4 Terminal (n=19, 25%)	EQ-5D-3L, EQVAS	EQ-5D score not presented for total sample, EQVAS 0.55 ± 0.22 (median 0.5)	The percentage of total sample who endorsed the worst response options of the EQ-5D increased with ALS severity across all dimensions Mean EQVAS scores decreased with increasing ALS severity from 0.74 to 0.37	Not available	54-80% of individuals with ALS reported moderate/severe problems in 4/5 EQ-5D dimensions (exception of 27% of people for *Anxiety/Depression*)	Not assessed – primary objective was not to assess psychometric properties of generic preference-based measure
Sherwood-Smith et al. (2000) [34]	USA	Cross-sectional study	To determine the concurrent validity of three self-administered HRQL questionnaires in patients with ALS	Outpatient clinic	*N* = 19	Age 60.5 (range 36-76), 42% Female, Time since diagnosis 20.5 months (range 2-62), ALSFRS-R score not presented, FVC 64% (range 17%-91%)	ALS severity not presented, Diagnosis classified according to the El Escorial criteria: Definite (n=9, 47%), Probable (n=4, 21%), Possible (n=5, 26%), Suspected (n=1, 5%)	QWB SA	QWB SA 0.43 (range 0-1)	Not available	QWB[a]: +SIP/ALS (r=0.55), +SF-36 (r=0.21)	Not available	Methodological quality[h]: adequate Rating[i]: Sufficient (inconsistent based on majority) Grading of quality of evidence[j]: very low

HRQL – Health-related quality of life, EQ-5D – EuroQol Five Dimension, EQVAS – EuroQol Visual Analogue Scale, ALSFRS-R – Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised, BDI – Beck Depression Inventory, ALSSS - Amyotrophic Lateral Sclerosis Severity Scale, MG - Myasthenia Gravis, FSHD - Facioscapulohumeral Muscular Dystrophy, MND – Motor Neuron Disease, ALSAQ-40 (ALS Assessment Questionnaire 40) subscales: ALSPM - Physical Mobility, ALSADL/I - Activities of Daily Living/Independence, ALSED - Eating and Drinking, ALSCOM –

Communication, ALSER - Emotional Reactions, ALS/HSS – ALS Health State Scale, FVC - forced vital capacity, QWB SA – Quality of Well-Being Self-Administered scale, SF-36 – Short form 36, SIP/ALS -19 – Sickness Impact Profile ALS-19

*Range of health utility scores from -0.306-0.885, with higher scores representing better health [42]
[a]Spearman's rank correlation coefficient
[b]Measures social support
[c]Proposed clinical stages developed by Jones et al. estimated using ALSFRS-R scores and modified King's ALS staging system to indicate ALS severity
[d]High severity: patients needed caregiver's assistance
[e]Low severity: patients did not need caregiver's assistance
[f]Pearson's product-moment correlation coefficient
[g]Correlation is significant at the 0.01 level (two-tailed)
[h]Determined using COSMIN's risk of bias checklist [27]
[i]Results rated against COSMIN's criteria for good measurement properties [26]: 50% of correlations (QWB: +SIP/ALS (r=0.55), +SF-36 (r=0.21)) in accordance with hypotheses, results rated as sufficient with an inconsistent rating from the majority of results
[j]Determined using the GRADE approach. More detail on how described in detail in the COSMIN manual [26]

Table 2. Convergent validity of the EQ-5D-3L and the QWB SA scale.

	Comparison Measure	Correlation (r)
EQ-5D-3L	EQVAS (Global Rating of Self-perceived Health) [33]	r=0.60
	ALSAQ-40 (Disease Specific Health Profile) [33] - PM - ADL/I - ED - COM - ER	 r=-0.60 r=-0.75 r=0.19 r=-0.32 r=-0.43
	ALSFRS-R (Disease Specific Functional Measure) [29]	r=0.72
	FSozU K-14 (Social Support) [29]	r=0.43
	BDI (Depression) [29]	r=-0.43
QWB -SA	SF-36 (Generic Health Profile) [34]	r=0.21
	SIP/ALS-19 (Disease Specific Health Profile) ([34]	r=0.55

EQ-5D-3L – EuroQol Five Dimension 3 Level, QWB SA – Quality of Well-Being Self-Administered scale, EQVAS – EuroQol Visual Analogue Scale, ALSAQ-40 - ALS Assessment Questionnaire 40; subscales: PM - Physical Mobility, ADL/I - Activities of Daily Living/Independence, ED - Eating and Drinking, COM - Communication, and ER - Emotional Reactions, ALSFRS-R – Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised, FSozU K-14 – Social Support, BDI – Beck Depression Inventory, SF-36 – Short form 36, SIP/ALS -19 – Sickness Impact Profile ALS-19

Chapter Three: Do Generic Preference-Based Measures Accurately Capture Areas of Health-Related Quality of Life Important to Individuals with Amyotrophic Lateral Sclerosis: A Content Validation Study

For submission to: *Quality of Life Research*

Do Generic Preference-Based Measures Accurately Capture Areas of Health-Related Quality of Life Important to Individuals with Amyotrophic Lateral Sclerosis: A Content Validation Study

Nicole Peters, BSc[1], Vanina Dal Bello-Haas, PT, PhD[1], Tara Packham, OTReg(Ont.), PhD[1], Jill Van Damme, MSc[1], Ayse Kuspinar, PT, PhD[1]

[1]*School of Rehabilitation Science, McMaster University, Hamilton, ON*

ORCID Numbers:

Nicole Peters: 0000-0001-6112-8436

Vanina Dal Bello-Haas: 0000-0003-4066-156X

Tara Packham: 0000-0002-5593-1975

Jill Van Damme: 0000-0002-6792-7447

Ayse Kuspinar: 0000-0002-3886-0407

Corresponding author:
Ayse Kuspinar, PT, PhD
Assistant Professor
School of Rehabilitation Science
McMaster University
1400 Main St. W. Room 435, IAHS
Hamilton, ON, L8S 1C7
Telephone: 905-525-9140 ext. 27823
kuspinaa@mcmaster.ca

Do Generic Preference-Based Measures Accurately Capture Areas of Health-Related Quality of Life Important to Individuals with Amyotrophic Lateral Sclerosis: A Content Validation Study

Abstract

Purpose: The objectives of this study were to: 1) assess the content validity of generic preference-based measures (GPBMs), and (2) examine the convergent validity of the EuroQol 5 Dimension 5 Level (EQ-5D-5L), against the Patient Generated Index (PGI) in Amyotrophic Lateral Sclerosis (ALS).

Methods: Participants were recruited from 3 clinical sites across Canada. The PGI, EQ-5D-5L and Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised (ALSFRS-R) were administered through an online or hardcopy survey and scores compared for convergent validation. Domains nominated by participants as important to their health-related quality of life were generated using the PGI, classified using the International Classification of Functioning, Disability and Health (ICF) and mapped onto GPBMs to determine content coverage.

Results: Fifty-two participants (N=28 female; 61.3 ± 11.6 mean age ± standard deviation (SD); 3.5 ± 2.9 mean ± SD years since diagnosis) completed this study. The top three ICF domains identified by participants were: *recreation and leisure*, *lower limb mobility*, and *interpersonal relationships*. The Quality of Well-Being Self-Administered (QWB-SA) scale had the highest content coverage (87%) and the Health Utilities Index 3 (HUI3) had the lowest (33%). Only two domains were covered by all GPBMs and no GPBM included all domains identified as important by participants. A moderate Pearson's correlation coefficient of 0.52 between the PGI and EQ-5D-5L was found.

Conclusion: The majority of GPBMs covered only approximately half of the domains important to individuals with ALS suggesting the need for an ALS specific preference-based measure to better reflect the health concerns of this population.

Key Words

Amyotrophic Lateral Sclerosis, Health-Related Quality of Life, Patient Reported Outcome Measures, Psychometric Properties, Economic Evaluation

Introduction

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease characterized by selective and progressive loss of voluntary motor neurons [1]. Individuals with ALS experience a range of symptoms related to the loss of muscle control in limb, bulbar and respiratory functions [2–4]. Consequently, activities of daily living, independence, and health-related quality of life (HRQL) are all impacted by the disease and as no curative treatment is currently available, optimal treatment is based on addressing symptoms and improving HRQL [2, 5–8].

Improvement in HRQL, "an individual's perception of how an illness and its treatment affect the physical, mental and social aspects of his or her life," [9, 10] is often considered the ultimate goal in healthcare [11]; and measures of HRQL can be used for treatment decision making and outcome evaluation purposes [11, 12]. Generic preference-based measures (GPBMs) are a type of HRQL measure designed to assess the cost-effectiveness of interventions due to their ability to produce a single index score, typically anchored from 0.0 (death) to 1.0 (perfect health) [13–15]. This value can be used to calculate quality-adjusted life years (QALYs) by capturing the effect of an intervention on one's quality of life (morbidity) and length of life (mortality) [13, 16]. GPBMs have been used to assess the HRQL of individuals with ALS and to aid researchers, policymakers and health care professionals in evaluating the cost-effectiveness of different treatment options [17].

While generic measures include a set of common domains relevant across a variety of health conditions, they may not capture all the domains that are impacted by specific health conditions. When this occurs, scores from GPBMs may be higher than the true impact, resulting in incorrect comparisons across interventions and populations [18].

Before a measure can be applied in practice, it must be tested to ensure that it is both reliable and valid [19, 20]. Content validity is the degree to which the content of an instrument accurately reflects the construct to be measured: a fundamental aspect in considering whether a measure can be used in a population [20]. However, the content validity of GPBMs in individuals with ALS has not yet been evaluated. Therefore, the primary objective of this study was to assess the content validity of GPBMs in ALS. The secondary objective of this study was to examine the convergent validity of the EuroQol 5 Dimension 5 Level (EQ-5D-5L), against the Patient Generated Index (PGI), in ALS.

Methods

Participants

Participants were recruited from the Canadian ALS Research Network (CALS) outpatient clinics across Western (Edmonton, AB), Central (Hamilton, ON), and Eastern (Fredericton, NB) Canada. Participants were eligible for inclusion if they were: 1) 18 years of age or older, 2) had a clinical diagnosis of ALS, and 3) able to communicate, verbally or electronically, in English. Individuals with severe frontotemporal dementia were excluded.

Outcome Measures

This study involved the administration of an online or hard copy (paper and pen) questionnaire consisting of sociodemographic and clinical information, the PGI, the EQ-5D-5L and the Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised (ALSFRS-R).

Sociodemographic and clinical information

Sociodemographic information consisting of age, sex, gender, residing region, highest education level, marital status, living situation and employment status was obtained. Clinical

information consisting of year of diagnosis, ALS symptom onset location, clinic location and clinic visitation frequency was recorded

Patient Generated Index (PGI)

The PGI [21] is an individualized measure previously utilized to identify areas of quality of life important to individuals in studies of content validity. It has been used with chronic conditions such as cancer [22–24], Parkinson's disease [25], Multiple Sclerosis (MS) [18] and Ankylosing Spondylitis [26]. It is completed in three stages. First, participants are asked to identify up to five of the most important areas of their lives affected by their health condition (i.e. ALS). Second, they are asked to rate the extent of impact of each area on their lives from 0 (the worst you could imagine) to 10 (exactly as you would like it to be). A supplementary sixth item is provided to rate all other areas of life not mentioned. This can include additional areas of life affected by the health condition, as well as non-health related areas. In the third phase, participants are asked to imagine that they could improve some or all their chosen areas. Participants are given twelve weighting points to distribute across the five potential areas they would like to have improved, as well as the sixth item indicative of all other areas not mentioned. They can distribute these weighting points in any manner they choose but cannot use more than 12 points in total. More points allocated to an area indicate greater importance and hope of improvement. An average of 2 or more weighting points per area is considered meaningful [25].

The rating and weighting points allocated to each area are then multiplied and summed to produce a single index score of overall HRQL from 0 to 10, with higher scores indicating better HRQL [21]. This score is typically reported as a percentage [27] and is intended to represent the extent to which reality matches expectations of perceived quality of life for those areas of life

patients most value an improvement [21, 23]. If there are missing data, an overall PGI score cannot be calculated.

EQ-5D-5L

The EQ-5D-5L [28], developed by the EuroQol Group, is a well-established and widely used GPBM of HRQL that consists of two parts [15, 29]. The first part (the descriptive system) assesses health in five domains: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each domain has five response levels, scored from 1 (no problems) to 5 (unable/extreme) [28]. A health utility value is derived from the five domains and is transformed into a single index score using a time-trade off (TTO) derived scoring system; a method of assigning values to health states from the population by asking respondents to choose between a shorter life in a state of perfect health or a longer life in a state of impaired health [30]. Canadian health utilities for the EQ-5D-5L range from -0.148 for the worst possible health state (55555; a score of 5 for each of the health domains) to 0.949 (11111; a score of 1 for each of the health domains) for the best possible health state [30]. The EQ-5D-5L describes a total of 3125 health states (5^5), has been translated into more than 170 languages world-wide, and takes only a few minutes for participants to complete [31]. The second part of the EQ-5D-5L consists of a visual analogue scale (VAS) of self-rated health, scored from 0 to 100 [31, 32]. The scores from the VAS cannot be used directly as weights in QALY calculations, as they not produce a single index value from a preference-based scoring system; however, the scores can be used as a subjective measure of one's self-perceived health [32].

Self-Administered ALSFRS-R

The ALSFRS-R [33] quantifies degree of functional impairment in ALS, and consists of 12 questions across 3 domains: bulbar, motor and respiratory. The questions are rated on a five-

point scale from 0 (complete dependence) to 4 (normal function) and a total ALSFRS-R score, ranging from 0 to 48, is produced through summation of the individual item scores; with higher scores indicating better health and increased predicted survival [33]. The self-administered version of the ALSFRS-R has demonstrated excellent reliability (intra-class correlation = 0.93, 95% CI0.88 to 0.96) and sensitivity to change over time [34].

More recently, studies have evaluated the measurement properties of the ALSFRS-R using Rasch analyses [35, 36] and longitudinal and survival analyses [37]. From their findings, researchers caution the reporting of a single total score and recommend reporting domain specific subscale scores organized into bulbar, motor and respiratory domains [33, 35–37]. A total subscale score for each domain is produced through summation of the corresponding items: items 1 to 3 (bulbar), items 4 to 9 (motor) and items 10 to 12 (respiratory) [35]. Bulbar and respiratory domains range in score from 0 to 12 whereby the motor domain ranges in score from 0 to 24; with higher scores indicating better function.

Procedure

Ethics approval for this cross-sectional study was obtained from McMaster University (HiREB #5664) and all sites in accordance with their respective research ethics boards. A designated clinician or research nurse located in clinic, identified and recruited eligible participants. If interested, participants were given a hard copy of information including the invitation letter, consent form, and survey. Alternatively, interested participants could provide their email and a link to complete the LimeSurvey [38] was then sent by the research team. Participants could complete the questionnaire package by themselves or with the aid of a caregiver.

The domains generated from the PGI were classified independently by two reviewers (NP and JVD) using the World Health Organization's International Classification of Function, Disability and Health (ICF) [39]. The ICF was used as it provided a framework for coding and a standardized description of health-related problems at various levels (impairments, activity limitations and participation restrictions). A third and fourth reviewer (AK and VDBH) were consulted if consensus was not reached or ambiguity in responses were present. A methodology was derived for the coding process to ensure accuracy of coding between reviewers and to capture all relevant domains:

1. The domain nominated by the participant was coded to the most specific ICF code; if the reported area covered more than one code, then all codes were assigned.
2. In order to eliminate subjectivity, reviewers coded all possible interpretations of the domains.
3. Broader categories were used for coding if there were inconsistencies between reviewers in order to be as conservative as possible.

All individual and overarching ICF domains were then mapped onto the seven leading GPBMs: the EQ-5D-5L [28], the SF-6D [40], the Health Utilities Index Mark 2 & 3 (HUI 2 & 3) [41], the Assessment of Quality of Life 8-Dimension (AQoL-8D) [42], the 15-Dimension (15D) [43] and the Quality of Well-Being Self-Administered (QWB-SA) scale [44]. Mapping was performed by two independent reviewers (NP and JVD), with a third reviewer (AK) for consultation, if needed. The content coverage of GPBMs was determined by the percentage of domains included in the GPBM that were also nominated by individuals with ALS using the PGI. High and low percentages correspond with high and low content coverage respectively.

This methodology, as outlined in Figure 1, followed that of similar studies assessing content validity of GPBMs using the PGI [18, 25, 45].

Sample Size

There are no sample size recommendations for content validation [46]. Therefore, our sample size calculation was based on: the number of participant responses needed to achieve data saturation (when no new relevant knowledge is being obtained from participants [47]) and; the recommended sample size for construct validation studies, which is a minimum of 50 patients total [20]. Studies have demonstrated that sample sizes around 15 to 20 are sufficient for saturation [48–50], therefore, to satisfy both recommendations we aimed to recruit between 45 to 60 participants across the 3 clinical sites.

Data Analysis

Descriptive statistics consisting of parametric measures such as mean, standard deviation, frequency and percentage were calculated to analyze participants' sociodemographic and clinical information. Scores for the PGI, EQ-5D-5L, and ALSFRS-R were computed according to their respective guidelines. Pearson's correlation coefficient was used to measure the strength of the association between PGI and EQ-5D-5L scores, as the data were normally distributed. Only complete data were used to assess association. Correlations with instruments measuring similar constructs should be greater than or equal to 0.50 [47], therefore to assess convergent validity, a correlation of at least 0.5 was hypothesized between the PGI and EQ-5D-5L [51].

Results

Table 1 presents the sociodemographic and clinical characteristics for the sample (N=52). A total of 35 participants completed the PGI in full. The mean age of the sample was 61 years

old and 54% were females. Approximately half of the sample (52%) completed the questionnaire in hard copy format, with the remaining completing the online format. Of the study participants, 67% completed the questionnaire without the assistance of a caregiver. Participants were distributed across Western (46%), Central (37%) and Eastern (17%) Canada, with 33% of individuals visiting their designated clinic every 3 months. The mean time since diagnosis was 3.5 (2.95) years. The time since diagnosis ranged from less than 1 year ago (27%) to five or more years (12%). For our sample, ALS symptoms first began to appear in the upper and/or lower limbs for 75% of the sample. The mean total ALSFRS-R score was 30.4 (9.4). The following subdomain scores were calculated for the ALSFRS-R: bulbar was 8.5 (3.5), scored out of 12; motor was 12.2 (6.2), scored out of 24; and respiratory was 9.8 (2.8), scored out of 12. The mean PGI score was 25.4 (14.1) and the mean EQ-5D-5L score was 0.55 (0.24).

Table 2 outlines the ICF domains identified by participants to be most affected by ALS and their frequency of appearance (n). There were 78 individual domains identified which resulted in 25 overarching ICF domains. The top 10 overarching domains identified were: *recreation and leisure* (17%), *lower limb mobility* (11%), *interpersonal relationships* (9%), *self-care* (7%), *housework and preparing meals* (6.5%), *speaking* (6%), *eating and swallowing* (5%), *work and employment* (4%), *upper limb mobility* (4%), and *daily routine and independence* (4%).

Figure 2 outlines the mean impact scores for each overarching ICF domain identified by participants. The three least impacted HRQL domains were *upper limb mobility* (mean score of 4.4), *self-care* (mean score of 3.6) and *lower limb mobility* (mean score of 3.2). The domains *work and employment* (mean score of 1.3) and *recreation and leisure* (mean score of 1.7) were the most impacted.

Figure 3 outlines the mean number of points (out of 12 points) that participants allocated to each of the overarching ICF domains. The most desired areas for improvement were *interpersonal relationships* (30% of points, mean of 3.0 points), *muscle & movement functions* (29% of points, mean of 2.9 points) and *speaking* (29% of points, mean of 2.9 points). The area with the least desire for improvement was *housework and preparing meals* (15% of points, mean of 1.5 points).

Table 3 presents the mapping of overarching ICF domains identified by individuals with ALS against GPBMs. The GPBM that covered the highest number of ICF domains identified by participants was the QWB-SA scale at 87% coverage. The HUI3 addressed the least number of domains with 33% coverage. The remaining GPBMs identified between 53-67% of domains. The domains covered by all GPBMs were *lower limb mobility* and *emotions*. The domains most commonly missing from GPBMs were: *structures involved in voice and speech*, which was only included in the QWB-SA scale; and *caring for household objects*, only included in the AQoL-8D. Figure 4 presents a scatter plot of EQ-5D-5L scores plotted against PGI scores. A positive moderate Pearson's correlation coefficient of 0.52 was observed between the two measures.

Discussion

This was the first study evaluating the content validity of GPBMs in individuals with ALS. Participants with ALS completed an individualized measure, the PGI, to evaluate the impact of the health condition on their HRQL. Commonly reported domains, identified as areas impacted by ALS and rated in terms of desire for improvement, were classified using the ICF and consequently mapped onto GPBMs to estimate the extent to which these generic measures captured domains that were important to individuals with ALS.

Individualized measures provide a standardized method to identify aspects of a health condition that impact patients' HRQL [52, 53]. The PGI [21] allowed individuals with ALS to identify the areas of their lives affected by their condition and assign a weight to each identified domain. The majority of GPBMs included approximately half of the areas reported on the PGI. The domains *lower limb mobility* and *emotions* were the only two areas identified by all GPBMs, however there was no one GPBM that included all the domains nominated by participants.

Domains self-identified as being affected by ALS encompassed three out of four ICF components - body structures (7%), body functions (13%), and activities and participation (80%). Domains nominated were relatively severely impacted, rated as very poor to poor [21] with an average score of 2.7 out of 10 (See Figure 2), and important to their quality of life with an average of 2.4 weighting points out of 12 (See Figure 3) allocated across domains for desire for improvement. For example, *recreation and leisure* was the most commonly reported domain and not only was it severely impacted by ALS but the desire for improvement in this domain was heavily weighted. This was true to some fashion for all domains nominated by participants as the average impact of ALS on domains was rated as severe. For example, it is well known that ALS progression results in a decline in physical health [2]. Studies have shown that HRQL is not necessarily dependent on patient's physical well-being but on their mental and social well-being [54–56]. The results of this study further demonstrate the impact of ALS on patients' social well-being and independence. Therefore, we would thus expect GPBMs used in ALS to capture these nominated domains, yet this was not the case.

In assessing the content coverage of GPBMs, the GPBM with the least coverage was the HUI3; evolved from the HUI/HUI1 and HUI2 [41]. The HUI3 has been widely used in clinical populations, including neurological conditions [57–59]. It includes eight HRQL domains that

focus on bodily functions: vision, speech, hearing, dexterity, ambulation, cognition, emotion and pain [60]. Only a third of the domains identified in our sample were covered by this measure i.e. *lower limb mobility*, *speaking*, *upper limb mobility*, *emotions* and *undertaking a task*. The HUI3 does not include domains relevant to social well-being, which was identified as important in our sample, and is missing many areas identified as important to their HRQL.

The QWB-SA scale had the highest percentage of included domains, with 87% of domains evaluated by the measure as relevant to individuals with ALS. The QWB-SA scale is a comprehensive measure of HRQL that combines scales of functioning with a measure of symptoms and problems [44]. However, the measure is very symptom and limitation focused, and did not seem to translate to our sample. ALS affects all areas of life including participation areas, and our study showed that the effects of the symptoms, rather than the symptoms themselves, were the most impacted in our sample. The QWB-SA scale did not capture this well. Furthermore, the administration of the QWB-SA scale is lengthy, i.e. takes around 14-minutes to complete [61], compared to the EQ-5D which only takes a few minutes to complete [31]; this may be one reason why it is not as widely used. Additionally, a study by Smith et al. (2000) [62] utilized the QWB-SA in ALS and found evidence corroborating decreased usage of the measure in this population, which may contribute further insight into the low utilization in ALS. Specifically, the measure demonstrated poor convergent validity with other measures of HRQL (i.e. the SF-36 and Sickness Impact Profile/ALS-19). The authors explained their results in relation to the general makeup of the test; with items and valuations that may not accurately capture the physical symptoms of individuals with ALS or provide equal weighting to items that are associated with ALS (such as self-care). Furthermore, in a systematic review [63] evaluating the psychometric properties of the QWB-SA scale in ALS, the measure was found to reveal

items that were not relevant to individuals with ALS and was found to be poor in quality: when assessed using COSMIN's modified Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach [47].

The EQ-5D-5L is a widely used GPBM and has previously been used in the ALS population [17, 64–68]. It covered 53% of domains identified by individuals with ALS (8/15 domains). However, it did not cover relevant and important domains such as *speaking*, *eating and swallowing* or *upper limb mobility,* which are affected by ALS. For instance, domains comprised by the EQ-5D-5L (Mobility, Self-care, Usual Activities, Pain/Discomfort and Anxiety/Depression) [28] are comparable to the impairments and activity limitations identified by our sample. Nonetheless, the social aspects of ALS identified by our sample as being impacted, were not explicitly addressed by the EQ-5D-5L as a distinctive domain; which resulted in a lower coverage than expected.

The mean EQ-5D-5L score in our sample indicates a moderate health state, whereas the mean PGI score indicates poor HRQL. A lower HRQL score on the PGI in comparison to the EQ-5D-5L, suggests that GPBMs may underestimate the effects of ALS on the HRQL of patients. Furthermore, a moderate Pearson's correlation between the PGI and the EQ-5D-5L was found in our study. A higher correlation was anticipated between the two HRQL measures [13], however it is not surprising considering there were identified domains not included in the EQ-5D-5L. Therefore, both the magnitude of correlation and percentage of content coverage in the EQ-5D-5L provide evidence of an overall lack of items relevant to individuals with ALS.

Strengths and Limitations

Our study is not without its strengths and limitations. To strengthen the generalizability of our study, participants were recruited from 3 different regions across Canada. Furthermore, we

had a fairly equal distribution of men and women in our sample. An additional strength is in the number of years since diagnosis in that it is reflective of the ALS population in large epidemiological studies [69]; with the majority of individuals diagnosed within two to five years. Moreover, in our sample 75% of participants indicated upper and/or lower symptoms as the primary symptom recognized at onset; and limb-onset ALS affects 65% to 75% of individuals diagnosed. Lastly, there was a wide range of ALSFRS-R scores in our sample, indicating mild to severe functional impairment, which is again reflective of the ALS population [69].

One limitation to this study was in utilizing the ICF as a coding framework. While the ICF has been used in similar studies, the framework is not all encompassing; therefore, some domains nominated by the sample, such as balance, were difficult to code. The second limitation was the amount of missing data present (N=17), particularly related to the completion of the hardcopy version of the PGI. As a result, we could only determine the magnitude of the association between the EQ-5D-5L and PGI for the 67% of the sample that had complete data.

Conclusion

Content of preference-based measures needs to be reflective of the population's values for accurate economic evaluation of treatments. Our results demonstrated that the majority of well-recognized GPBMs included only approximately half of the domains important to those living with ALS. The most commonly used GPBM in ALS, the EQ-5D-5L, correlated moderately with the PGI, however it underestimated the impact of ALS on the HRQL of patients. Likewise, in assessing the content validity of GPBMs, there were domains that were not identified, or that were inaccurately represented and not relevant to our sample. Findings from this study suggest the need for the development of an ALS specific preference-based measure

with items that will capture the areas of life important to people with ALS and provide population-specific values that can be utilized for the assessment of treatment implications.

Declarations

Acknowledgments. The authors would like to thank the neurologists at each clinical site for their assistance with recruitment of participants.

Funding. This work was supported by the 2018 ALS Project Grant awarded by the ALS Society of Canada.

Authors' contributions. All authors (NP, VDBH, TP, JVD, AK) contributed to the study conception and design. Data collection and analysis were completed by NP, in consultation with study co-authors. Classification and mapping of domains were performed by NP and JVD. The first draft of the manuscript was written by NP. Preliminary edits were made by AK and all authors commented on previous iterations of the manuscript. All authors read and approved the final manuscript and NP is the guarantor of the paper.

Data availability. The data that supports the findings of this study are available in this published article and from the corresponding author (AK) on reasonable request.

Compliance with ethical standards

Conflicts of interest. The authors declare that they have no conflict of interest.

Ethics Approval. This study was approved by the ethics committee of McMaster University (HiREB #5664) and all sites in accordance with their respective research ethics boards. The

study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Consent to participate. Informed consent was obtained from all individual participants included in the study.

References

1. Rowland, L. P., & Shneider, N. A. (2001). Amyotrophic Lateral Sclerosis. *New England Journal of Medicine*, *344*(22), 1688–1700. https://doi.org/10.1056/NEJM200105313442207
2. Zarei, S., Carr, K., Reiley, L., Diaz, K., Guerra, O., Altamirano, P. F., … Chinea, A. (2015). A comprehensive review of amyotrophic lateral sclerosis. *Surgical Neurology International*, *6*, 171. https://doi.org/10.4103/2152-7806.169561
3. Pasinelli, P., & Brown, R. H. (2006). Molecular biology of amyotrophic lateral sclerosis: insights from genetics. *Nature Reviews Neuroscience*, *7*(9), 710–723. https://doi.org/10.1038/nrn1971
4. Andersen, P. M., Borasio, G. D., Dengler, R., Hardiman, O., Kollewe, K., Leigh, P. N., … Tomik, B. (2005). EFNS task force on management of amyotrophic lateral sclerosis: guidelines for diagnosing and clinical care of patients and relatives. An evidence-based review with good practice points. *European Journal of Neurology*, *12*(12), 921–938. https://doi.org/10.1111/j.1468-1331.2005.01351.x
5. Simmons, Z. (2015). Patient-Perceived Outcomes and Quality of Life in ALS. *Neurotherapeutics*, *12*(2), 394–402. https://doi.org/10.1007/s13311-014-0322-x
6. Handy, C. R., Krudy, C., Boulis, N., & Federici, T. (2011). Pain in Amyotrophic Lateral Sclerosis: A Neglected Aspect of Disease. *Neurology Research International*, *2011*, 1–8. https://doi.org/10.1155/2011/403808
7. Swash, M. (1997). Health outcome and quality-of-life measurements in amyotrophic lateral sclerosis. *Journal of Neurology*, *244*(S2), S26–S29. https://doi.org/10.1007/BF03160578

8. Bensimon, G., Lacomblez, L., & Meininger, V. (1994). A Controlled Trial of Riluzole in Amyotrophic Lateral Sclerosis. *New England Journal of Medicine*, *330*(9), 585–591. https://doi.org/10.1056/NEJM199403033300901
9. Mayo, N. E. (2015). *ISOQOL Dictionary of quality of life and health outcomes measurement*. ISOQOL. Retrieved from https://books.google.ca/books?id=cKjksgEACAAJ
10. Karimi, M., & Brazier, J. (2016). Health, Health-Related Quality of Life, and Quality of Life: What is the Difference? *PharmacoEconomics*, *34*(7), 645–649. https://doi.org/10.1007/s40273-016-0389-9
11. Ruta, D. A., & Garratt, A. M. (2013). Health status to quality of life measurement. In C. Jenkinson (Ed.), *Measuring health and medical outcomes* (pp. 138–159). London; New York: Routledge. Retrieved from https://www.taylorfrancis.com/books/e/9781315072159
12. Guyatt, G. H. (1993). Measuring Health-Related Quality of Life. *Annals of Internal Medicine*, *118*(8), 622. https://doi.org/10.7326/0003-4819-118-8-199304150-00009
13. Brazier, J., Ratcliffe, J., Salomon, J. A., & Tsuchiya, A. (2017). *Measuring and valuing health benefits for economic evaluation* (Second edition.). Oxford: Oxford University Press.
14. Torrance, G. W. (1987). Utility approach to measuring health-related quality of life. *Journal of Chronic Diseases*, *40*(6), 593–600. https://doi.org/10.1016/0021-9681(87)90019-1
15. Brauer, C. A., Rosen, A. B., Greenberg, D., & Neumann, P. J. (2006). Trends in the Measurement of Health Utilities in Published Cost-Utility Analyses. *Value in Health*, *9*(4), 213–218. https://doi.org/10.1111/j.1524-4733.2006.00116.x

16. Neumann, P. J., Goldie, S. J., & Weinstein, M. C. (2000). Preference-Based Measures in Economic Evaluation in Health Care. *Annual Review of Public Health*, *21*(1), 587–611. https://doi.org/10.1146/annurev.publhealth.21.1.587

17. Lopez-Bastida J., Perestelo-Perez L., Monton-Alvarez F., Serrano-Aguilar P., & Alfonso-Sanchez J.L. (2009). Social economic costs and health-related quality of life in patients with amyotrophic lateral sclerosis in Spain. *Amyotrophic Lateral Sclerosis*, *10*(4), 237–243. https://doi.org/10.1080/17482960802430781

18. Kuspinar, A., & Mayo, N. E. (2013). Do generic utility measures capture what is important to the quality of life of people with multiple sclerosis? *Health and Quality of Life Outcomes*, *11*(1), 71. https://doi.org/10.1186/1477-7525-11-71

19. Souza, A. C. de, Alexandre, N. M. C., Guirardello, E. de B., Souza, A. C. de, Alexandre, N. M. C., & Guirardello, E. de B. (2017). Psychometric properties in instruments evaluation of reliability and validity. *Epidemiologia e Serviços de Saúde*, *26*(3), 649–659. https://doi.org/10.5123/S1679-49742017000300022

20. de Vet, H. C. W. (Ed.). (2011). *Measurement in medicine: a practical guide*. Cambridge ; New York: Cambridge University Press.

21. Ruta, D. A., Garratt, A. M., Leng, M., Russell, I. T., & MacDonald, L. M. (1994). A New Approach to the Measurement of Quality of Life: The Patient-Generated Index. *Medical Care*, *32*(11), 1109–1126. https://doi.org/10.1097/00005650-199411000-00004

22. Tavernier, S. S., Totten, A. M., & Beck, S. L. (2011). Assessing Content Validity of the Patient Generated Index Using Cognitive Interviews. *Qualitative Health Research*, *21*(12), 1729–1738. https://doi.org/10.1177/1049732311420169

23. Aburub, A. S., Gagnon, B., Rodríguez, A. M., & Mayo, N. E. (2016). Using a personalized measure (Patient Generated Index (PGI)) to identify what matters to people with cancer. *Supportive Care in Cancer*, *24*(1), 437–445. https://doi.org/10.1007/s00520-015-2821-7

24. Tavernier, S. S., Beck, S. L., Clayton, M. F., Pett, M. A., & Berry, D. L. (2011). Validity of the Patient Generated Index as a Quality-of-Life Measure in Radiation Oncology. *Oncology Nursing Forum*, *38*(3), 319–329. https://doi.org/10.1188/11.ONF.319-329

25. Kuspinar, A., Mate, K., Lafontaine, A.-L., & Mayo, N. (2019). Evaluating the content validity of generic preference-based measures for use in Parkinson's disease. *Parkinsonism & Related Disorders*, *62*, 112–116. https://doi.org/10.1016/j.parkreldis.2019.01.014

26. Haywood, K. L., Garratt, A. M., Dziedzic, K., & Dawes, P. T. (2003). Patient centered assessment of ankylosing spondylitis-specific health related quality of life: evaluation of the Patient Generated Index. *The Journal of Rheumatology*, *30*(4), 764–773.

27. Martin, F., Camfield, L., Rodham, K., Kliempt, P., & Ruta, D. (2007). Twelve years–experience with the Patient Generated Index (PGI) of quality of life: a graded structured review. *Quality of Life Research*, *16*(4), 705–715. https://doi.org/10.1007/s11136-006-9152-6

28. Herdman, M., Gudex, C., Lloyd, A., Janssen, Mf., Kind, P., Parkin, D., … Badia, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*, *20*(10), 1727–1736. https://doi.org/10.1007/s11136-011-9903-x

29. Brazier, J., Ara, R., Rowen, D., & Chevrou-Severac, H. (2017). A Review of Generic Preference-Based Measures for Use in Cost-Effectiveness Models. *PharmacoEconomics*, *35*(S1), 21–31. https://doi.org/10.1007/s40273-017-0545-x

30. Xie, F., Pullenayegum, E., Gaebel, K., Bansback, N., Bryan, S., Ohinmaa, A., … Canadian EQ-5D-5L Valuation Study Group. (2016). A Time Trade-off-derived Value Set of the EQ-5D-5L for Canada. *Medical Care*, *54*(1), 98–105. https://doi.org/10.1097/MLR.0000000000000447

31. Reenen, M. van, & Janssen, B. (2019). EQ-5D-5L User Guide: Basic information on how to use the EQ-5D-5L instrument. EuroQol Research Foundation. Retrieved from https://euroqol.org/publications/user-guides

32. Brooks, R. (1996). EuroQol: the current state of play. *Health Policy*, *37*(1), 53–72. https://doi.org/10.1016/0168-8510(96)00822-6

33. Cedarbaum, J. M., Stambler, N., Malta, E., Fuller, C., Hilt, D., Thurmond, B., & Nakanishi, A. (1999). The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. *Journal of the Neurological Sciences*, *169*(1–2), 13–21. https://doi.org/10.1016/S0022-510X(99)00210-5

34. Montes, J., Levy, G., Albert, S., Kaufmann, P., Buchsbaum, R., Gordon, P. H., & Mitsumoto, H. (2006). Development and evaluation of self-administered version of the ALSFRS-R. *Neurology*, *67*, 1294–1296. https://doi.org/10.1212/01.wnl.0000238505.22066.fc

35. Franchignoni, F., Mora, G., Giordano, A., Volanti, P., & Chiò, A. (2013). Evidence of multidimensionality in the ALSFRS-R Scale: a critical appraisal on its measurement properties using Rasch analysis. *Journal of Neurology, Neurosurgery & Psychiatry*, *84*(12), 1340–1345. https://doi.org/10.1136/jnnp-2012-304701

36. Franchignoni, F., Mandrioli, J., Giordano, A., Ferro, S., & ERRALS Group. (2015). A further Rasch study confirms that ALSFRS-R does not conform to fundamental

measurement requirements. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *16*(5–6), 331–337. https://doi.org/10.3109/21678421.2015.1026829

37. Rooney, J., Burke, T., Vajda, A., Heverin, M., & Hardiman, O. (2017). What does the ALSFRS-R really measure? A longitudinal and survival analysis of functional dimension subscores in amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, *88*(5), 381–385. https://doi.org/10.1136/jnnp-2016-314661
38. Limesurvey GmbH., & Schmitz, C. (2020). *LimeSurvey: An Open Source survey tool*. Hamburg, Germany: Limesurvey GmbH. Retrieved from http://www.limesurvey.org
39. World Health Organization (Ed.). (2001). *International classification of functioning, disability and health: ICF*. Geneva: World Health Organization.
40. Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*, *21*(2), 271–292. https://doi.org/10.1016/S0167-6296(01)00130-8
41. Horsman, J., Furlong, W., Feeny, D., & Torrance, G. (2003). The Health Utilities Index (HUI®): concepts, measurement properties and applications. *Health and Quality of Life Outcomes*, *1*(1), 54. https://doi.org/10.1186/1477-7525-1-54
42. Richardson, J., Monash University, Centre for Health Economics, Monash University, & Faculty of Business and Economics. (2011). *Modelling the utility of health states with the assessment of quality of life (AQoL) 8D instrument: overview and utility scoring algorithm*. Clayton, Victoria, Australia: Monash University, Business and Economics, Centre for Health Economics. Retrieved from http://www.buseco.monash.edu.au/centres/che/pubs/researchpaper63.pdf

43. Sintonen, H. (2001). The 15D instrument of health-related quality of life: properties and applications. *Annals of Medicine*, *33*(5), 328–336. https://doi.org/10.3109/07853890109002086

44. Seiber, W. J., Groessl, E. J., David, K. M., Ganiats, T. G., & Kaplan, R. M. (2008). Quality of Well-Being Self-Administered Scale (QWB-SA). San Diego Health Services Research Centre, University of California. Retrieved from https://www.researchgate.net/profile/Kristin_Kistler/publication/252316672_Quality_of_well_being_self-administered_QWB-SA_scale/links/5437d6990cf2590375c55a65/Quality-of-well-being-self-administered-QWB-SA-scale.pdf

45. Mayo, N. E., Moriello, C., Asano, M., van der Spuy, S., & Finch, L. (2011). The extent to which common health-related quality of life indices capture constructs beyond symptoms and function. *Quality of Life Research*, *20*(5), 621–627. https://doi.org/10.1007/s11136-010-9801-7

46. Mason, M. (2010). Sample Size and Saturation in PhD Studies Using Qualitative Interviews. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *Vol 11*, No 3 (2010): Methods for Qualitative Management Research in the Context of Social Systems Thinking. https://doi.org/10.17169/FQS-11.3.1428

47. Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, *27*(5), 1147–1157. https://doi.org/10.1007/s11136-018-1798-3

48. Bertaux, D. (Ed.). (1981). *Biography and society: the life history approach in the social sciences*. Beverly Hills, Calif: Sage Publications.

49. Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods*, *18*(1), 59–82. https://doi.org/10.1177/1525822X05279903

50. Green, J., & Thorogood, N. (2009). *Qualitative methods for health research* (2. ed.). London: SAGE.

51. Mukaka, M. M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal: The Journal of Medical Association of Malawi*, *24*(3), 69–71.

52. Dijkers, M. P. (2003). Individualization in quality of life measurement: Instruments and approaches. *Archives of Physical Medicine and Rehabilitation*, *84*, S3–S14. https://doi.org/10.1053/apmr.2003.50241

53. Joyce, C. R. B., McGee, H. M., & O̊Boyle, C. A. (1999). *Individual quality of life: approaches to conceptualisation and assessment*. Amsterdam: Harwood Academic Publishers. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=526533

54. Chio, A. (2004). A cross sectional study on determinants of quality of life in ALS. *Journal of Neurology, Neurosurgery & Psychiatry*, *75*(11), 1597–1601. https://doi.org/10.1136/jnnp.2003.033100

55. O'Doherty, L. J., Hickey, A., & Hardiman, O. (2010). Measuring life quality, physical function and psychological well-being in neurological illness. *Amyotrophic Lateral Sclerosis*, *11*(5), 461–468. https://doi.org/10.3109/17482960903552488

56. Simmons, Z., Bremer, B. A., Robbins, R. A., Walsh, S. M., & Fischer, S. (2000). Quality of life in ALS depends on factors other than strength and physical function. *Neurology*, *55*(3), 388–392. https://doi.org/10.1212/wnl.55.3.388

57. Fisk, J. D. (2005). A comparison of health utility measures for the evaluation of multiple sclerosis treatments. *Journal of Neurology, Neurosurgery & Psychiatry*, *76*(1), 58–63. https://doi.org/10.1136/jnnp.2003.017897

58. Pohar, S. L., & Allyson Jones, C. (2009). The burden of Parkinson disease (PD) and concomitant comorbidities. *Archives of Gerontology and Geriatrics*, *49*(2), 317–321. https://doi.org/10.1016/j.archger.2008.11.006

59. Edwards, J. D., Koehoorn, M., Boyd, L. A., & Levy, A. R. (2010). Is Health-Related Quality of Life Improving After Stroke?: A Comparison of Health Utilities Indices Among Canadians With Stroke Between 1996 and 2005. *Stroke*, *41*(5), 996–1000. https://doi.org/10.1161/STROKEAHA.109.576678

60. Feeny, D., Furlong, W., Boyle, M., & Torrance, G. W. (1995). Multi-Attribute Health Status Classification Systems: Health Utilities Index. *PharmacoEconomics*, *7*(6), 490–502. https://doi.org/10.2165/00019053-199507060-00004

61. Andresen, E. M., Rothenberg, B. M., & Kaplan, R. M. (1998). Performance of a self-administered mailed version of the Quality of Well-Being (QWB-SA) questionnaire among older adults. *Medical Care*, *36*(9), 1349–1360. https://doi.org/10.1097/00005650-199809000-00007

62. Smith, P. S., Crossley, B., Greenberg, J., Wilder, C., & Carroll, B. (2000). Agreement among three quality of life measures in patients with ALS. *Amyotrophic lateral sclerosis*

and other motor neuron disorders : official publication of the World Federation of Neurology, Research Group on Motor Neuron Diseases, *1*(4), 269–75.

63. Peters, N., Dal Bello-Haas, V., Packham, T., Mehdipour, A., & Kuspinar, A. Psychometric Properties of Preference-Based Measures for Economic Evaluation in Amyotrophic Lateral Sclerosis: A Systematic Review. *Submitted 2020 to Quality of Life Research*
64. Winter Y., Schepelmann K., Spottke A.E., Claus D., Grothe C., Schroder R., … Dodel R. (2010). Health-related quality of life in ALS, myasthenia gravis and facioscapulohumeral muscular dystrophy. *Journal of Neurology*, *257*(9), 1473–1481. https://doi.org/10.1007/s00415-010-5549-9
65. Jones A.R., Jivraj N., Balendra R., Murphy C., Kelly J., Thornhill M., … Al-Chalabi A. (2014). Health utility decreases with increasing clinical stage in amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *15*(3–4), 285–291. https://doi.org/10.3109/21678421.2013.872149
66. Ilse, B., Prell, T., Walther, M., Hartung, V., Penzlin, S., Tietz, F., … Grosskreutz, J. (2015). Relationships between disease severity, social support and health-related quality of life in patients with amyotrophic lateral sclerosis. *Social Indicators Research*, *120*(3), 871–882. https://doi.org/10.1007/s11205-014-0621-y
67. Green C., Kiebert G., Murphy C., Mitchell J.D., O'Brien M., Burrell A., & Leigh P.N. (2003). Patients' health-related quality-of-life and health state values for motor neurone disease/amyotrophic lateral sclerosis. *Quality of Life Research*, *12*(5), 565–574. https://doi.org/10.1023/A:1025052609818
68. Kiebert G.M., Green C., Murphy C., Mitchell J.D., O'Brien M., Burrell A., & Leigh P.N. (2001). Patients' health-related quality of life and utilities associated with different stages

of amyotrophic lateral sclerosis. *Journal of the Neurological Sciences*, *191*(1–2), 87–93. https://doi.org/10.1016/S0022-510X%2801%2900616-5

69. Longinetti, E., & Fang, F. (2019). Epidemiology of amyotrophic lateral sclerosis: an update of recent literature. *Current Opinion in Neurology*, *32*(5), 771–776. https://doi.org/10.1097/WCO.0000000000000730

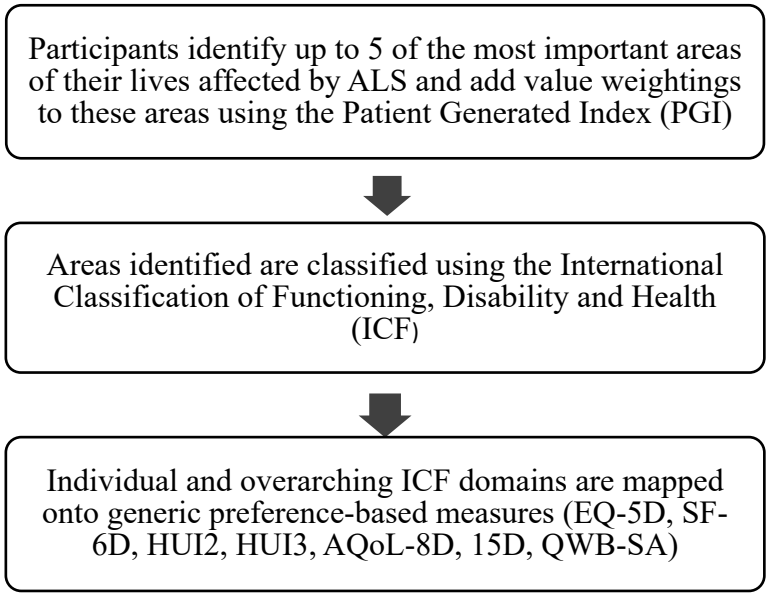


Figure 1. Outline of the study procedures.

Table 1. Sociodemographic and clinical characteristics of sample (N=52).

Characteristics	**Mean (SD) or N (%)**
Age (years), range	61.3 (11.6), 23-86
Females	28 (53.8)
Residing Region	
Western Canada	24 (46.2)
Central Canada	19 (36.5)
Eastern Canada	9 (17.3)
Education (highest level)	
Less than high school	3 (5.8)
High School	18 (34.6)
CEGEP/College	22 (42.3)
Bachelor's	7 (13.5)
Graduate	2 (3.8)
Marital status	
Married/Common Law	39 (75.0)
Divorced/Separated	6 (11.5)
Widowed	3 (5.8)
Never married	4 (7.7)
Living situation	
Own home	49 (94.2)
Retirement facility	3 (5.8)
Employment status*	
Full-time	4 (7.8)
Self-employed	2 (3.9)
Short-term disability	1 (1.9)
Long-term disability	15 (29.4)
Retired	23 (44.2)
Unemployed	4 (7.8)
Other[a]	2 (3.9)
ALS clinic attendance	
Western Canada	
- Edmonton, AB	23 (44.2)
Central Canada	
- Hamilton, ON	19 (36.5)
Eastern Canada	
- Fredericton, NB	9 (17.3)
- Halifax, NS	1 (1.9)

Clinic visitation frequency	
Monthly	4 (7.7)
3-5 months	35 (67.3)
6 months	6 (11.5)
Yearly	2 (3.8)
Other[b]	5 (9.6)
Years since diagnosis	
Mean, range	3.5 (2.95), <1-14
≤ 1	14 (26.9)
2	15 (28.8)
3	10 (19.2)
4	7 (13.5)
≥ 5	6 (11.5)
Location of symptom onset	
Upper limb	18 (34.6)
Lower limb	18 (34.6)
Neck/face	12 (23.1)
Breathing	1 (1.9)
Upper and lower limb	3 (5.8)
PGI** [0-100]	25.4 (14.1)
EQ-5D-5L*** [0-1]	0.55 (0.24)
EQVAS*** [0-100]	54.6 (21.0)
ALSFRS-R	
Total score*** [0-48]	30.4 (9.4)
Subdomain score	
- Bulbar**** [0-12]	8.5 (3.5)
- Motor*** [0-24]	12.2 (6.2)
- Respiratory*** [0-12]	9.8 (2.8)

SD – standard deviation, N – sample size, % - frequency, PGI – Patient Generated Index, EQ-5D-5L – EuroQol 5 Dimension 5 Level, EQVAS – EuroQol Visual Analogue Scale, ALSFRS-R – Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised

[a]Short term sick leave (1(1.9)); Currently not working (1(1.9))

[b]Just diagnosed (4(7.7)); 3 months but more frequently for trials (1(1.9))

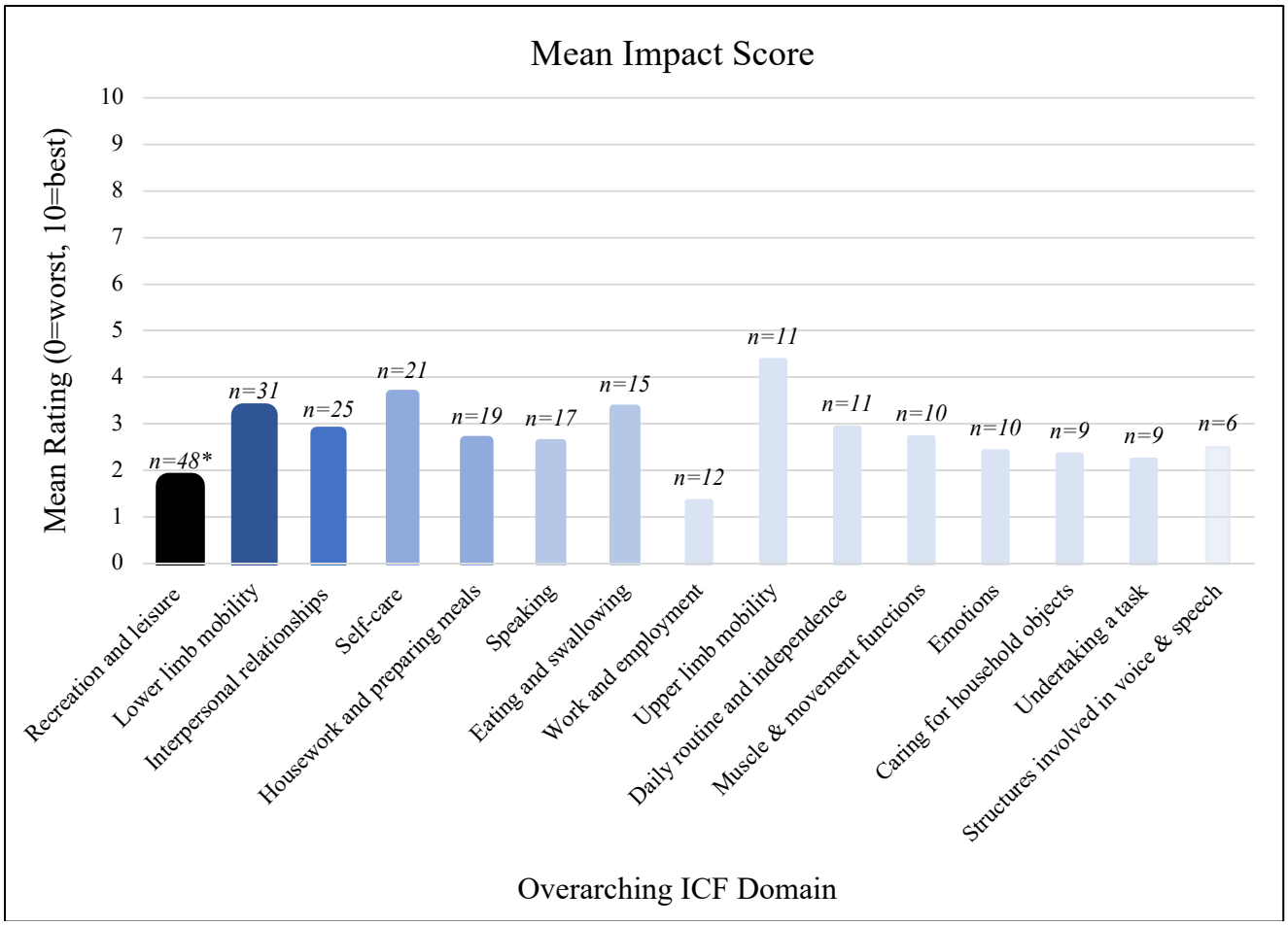
*Missing data (N=1) **Missing data (N=17) ***Missing data (N=6) ****Missing data (N=5)

Table 2. ICF domains identified by participants using the Patient Generated Index, organized from most frequent to least frequent (n=15).

ICF Domain	ICF Code	Frequency n (%)	Overarching ICF Domain	Overarching ICF Code	Frequency n (%)
Recreation and leisure	d920	27 (9.3)	Recreation and leisure	d920	49 (16.8)
Socializing	d9205	12 (4.1)			
Hobbies	d9204	2 (0.7)			
Crafts	d9203	1 (0.3)			
Sports	d9201	2 (0.7)			
Play	d9200	5 (1.7)			
Mobility, unspecified	d499	6 (2.1)	Lower limb mobility	d450-d469; d475; d499	31 (10.7)
Driving	d475	6 (2.1)			
Moving around using equipment	d465	2 (0.7)			
Moving around in different locations	d460	3 (1.0)			
Climbing	d4551	1 (0.3)			
Walking	d450	13 (4.5)			
Family relationships	d760	13 (4.5)	Interpersonal relationships	d7	25 (8.6)
Parent-child relationships	d7600	2 (0.7)			
Sexual relationships	d7702	4 (1.4)			
Spousal relationships	d7701	2 (0.7)			
Physical contact in relationships	d7105	3 (1.0)			
Basic interpersonal interactions	d710	1 (0.3)			
Self-care, unspecified	d599	4 (1.4)	Self-care	d5	21 (7.2)
Looking after one's health, other specified	d5708	1 (0.3)			
Dressing	d540	6 (2.1)			
Toileting	d530	4 (1.4)			
Caring for hair	d5202	2 (0.7)			
Caring for skin	d5200	1 (0.3)			
Washing oneself	d510	3 (1.0)			
Household tasks, other specified and unspecified	d649	1 (0.3)	Housework and preparing meals	d630-d649	19 (6.5)
Doing housework	d640	7 (2.4)			
Preparing meals	d630	11 (3.8)			

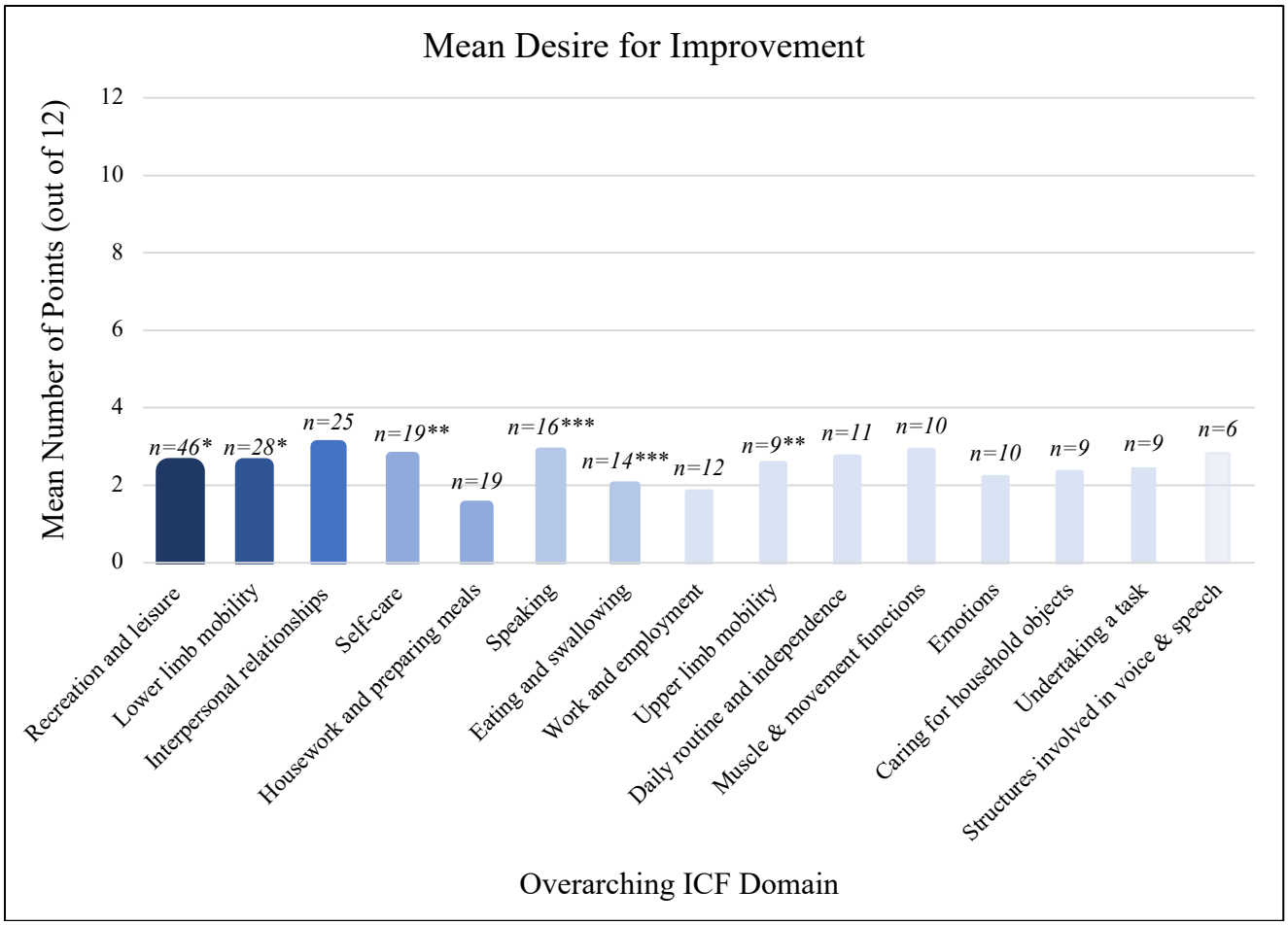
Communication, producing, other specified and unspecified	d349	2 (0.7)	Speaking	d330-d349	17 (5.8)
Speaking	d330	15 (5.2)			
Eating	d550	12 (4.1)	Eating and swallowing	d550; b5105	15 (5.2)
Swallowing	b5105	3 (1.0)			
Remunerative employment	d850	12 (4.1)	Work and employment	d850	12 (4.1)
Hand and arm use	d445	2 (0.7)	Upper limb mobility	d430-d449	11 (3.8)
Fine hand use	d440	8 (2.7)			
Lifting	d4300	1 (0.3)			
Carrying out daily routine	d230	11 (3.8)	Daily routine and independence	d230	11 (3.8)
Supportive functions of arm or leg	b7603	3 (1.0)	Muscle & movement functions	b730-b749; b750-b789	10 (3.4)
Control of voluntary movement functions	b760	2 (0.7)			
Involuntary movement of reaction function	b755	2 (0.7)			
Muscle endurance functions	b740	1 (0.3)			
Muscle power functions	b730	2 (0.7)			
Emotional functions, other specified	b1528	8 (2.7)	Emotions	b152	10 (3.4)
Emotional functions	b152	2 (0.7)			
Caring for household objects, specified	d6508	1 (0.3)	Caring for household objects	d650	9 (3.1)
Taking care of animals	d6506	2 (0.7)			
Taking care of plants, indoors and outdoors	d6505	3 (1.0)			
Caring for household objects	d650	3 (1.0)			
Undertaking a single task independently	d2102	8 (2.7)	Undertaking a task	d210	9 (3.1)
Undertaking a simple task	d2100	1 (0.3)			
Structure of pharynx, other specified	s3308	2 (0.7)	Structures involved in voice & speech	s3	6 (2.1)
Structure of lips	s3204	2 (0.7)			
Tongue	s3203	2 (0.7)			

n=frequency of appearance



n=frequency of appearance
*Missing data (N=1)

Figure 2. Distribution of mean impact scores, from 0 (the worst one could imagine) to 10 (exactly as one would like it to be), for the overarching ICF domains identified by the sample.



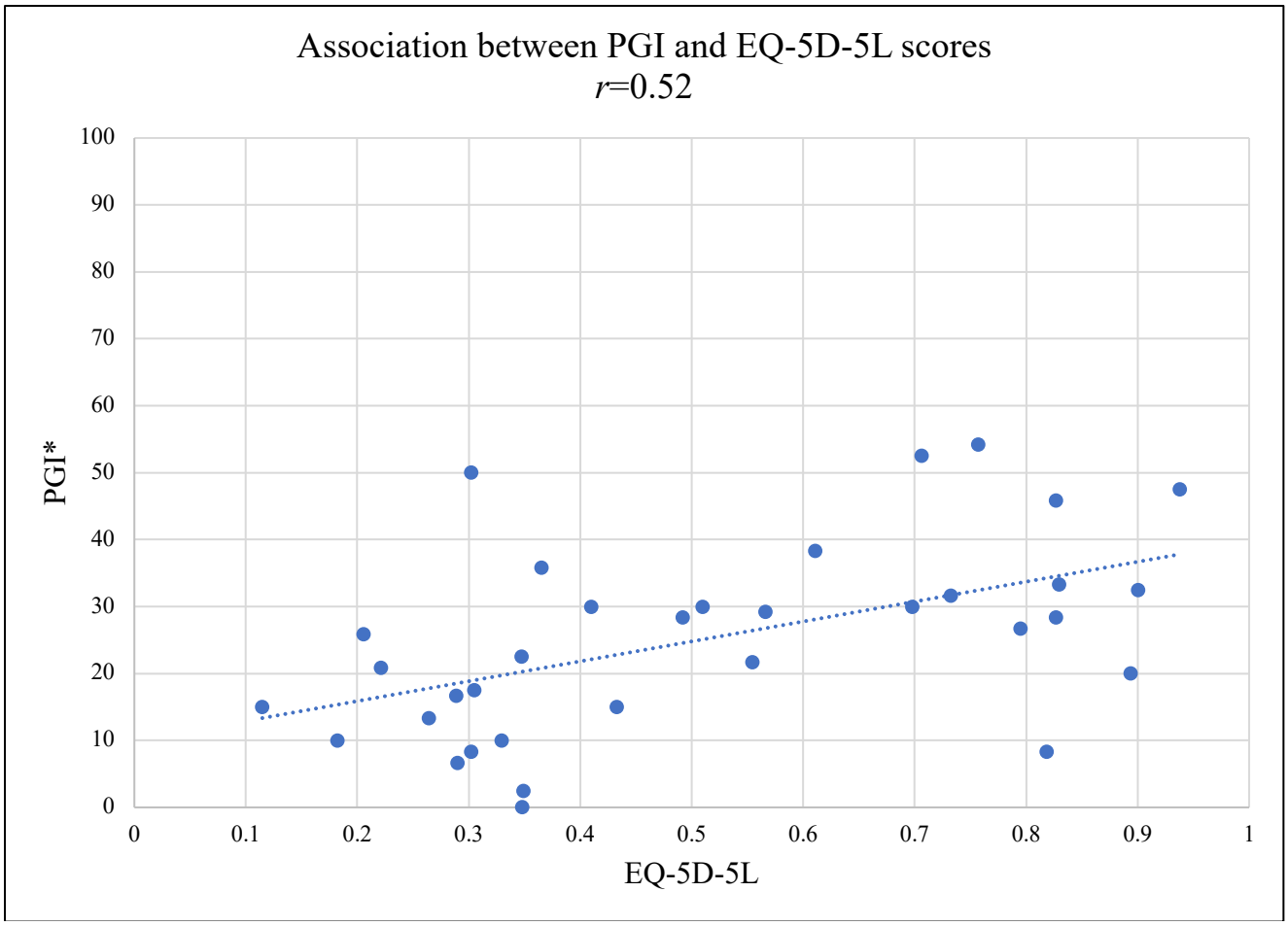
n=frequency of appearance
*Missing data (N=3) **Missing data (N=2) ***Missing data (N=1)

Figure 3. Mean desire for improvement displayed as a distribution of mean number of points (out of 12 points, where higher points indicate a greater importance for improvement) allocated to classified overarching ICF domains.

Table 3. Overarching ICF domains identified by sample mapped onto generic preference-based measures (n=15).

Overarching ICF Domain	Generic Preference-Based Measure						
	EQ-5D	**SF-6D**	**HUI2**	**HUI3**	**AQoL-8D**	**15D**	**QWBSA**
Recreation and leisure	Y	Y	N	N	Y	Y	Y
Lower limb mobility	Y	Y	Y	Y	Y	Y	Y
Interpersonal relationships	Y	Y	N	N	Y	Y	Y
Self-care	Y	Y	Y	N	Y	Y	Y
Housework and preparing meals	Y	Y	N	N	Y	Y	Y
Speaking	N	N	Y	Y	Y	Y	Y
Eating and swallowing	N	N	Y	N	Y	Y	Y
Work and employment	Y	Y	N	N	N	Y	Y
Upper limb mobility	N	Y	Y	Y	N	N	Y
Daily routine and independence	Y	Y	Y	N	N	Y	Y
Muscle & movement functions	N	N	Y	N	N	N	Y
Emotions	Y	Y	Y	Y	Y	Y	Y
Caring for household objects	N	N	N	N	Y	N	N
Undertaking a task	N	N	Y	Y	Y	N	N
Structures involved in voice & speech	N	N	N	N	N	N	Y
% Yes	53%	60%	60%	33%	67%	67%	87%

EQ-5D-5L – EuroQol 5 Dimension 5 Level, HUI2 – Health Utilities Index 2, HUI3 – Health Utilities Index 3, AQoL-8D - Assessment of Quality of Life 8-Dimension, 15D – 15 Dimension, QWB-SA - Quality of Well-Being Self-Administered scale, Y – Yes; covered by the generic preference-based measure, N – No; not covered by the generic preference-based measure



PGI – Patient Generated Index, EQ-5D-5L – EuroQol 5 Dimension 5 Level
*Missing data (N=17)

Figure 4. Scatter plot displaying the relationship between PGI scores and EQ-5D-5L scores.

Chapter Four: Discussion

This thesis contributes evidence towards the psychometric properties of generic preference-based measures in ALS. Two novel studies were conducted to: 1) summarize which psychometric properties have been reported in the literature; and, 2) evaluate the content validity of GPBMs in ALS. This chapter provides a review of the studies conducted and an interpretation of the combined results. The importance of the work, potential limitations and future directions are explored, along with the conclusions that can be drawn based on the current results.

4.1 Overview of Thesis Results

In Chapter Two, a systematic review was conducted to summarize the psychometric properties of a specific type of HRQL measure, GPBMs, in ALS. Our results showed that the EQ-5D-3L was the most commonly used GPBM in the ALS literature. Although the EQ-5D-3L demonstrated convergent and known-groups validity in ALS, significant floor effects were observed with the individual items.

In Chapter Three, a content validation study of GPBMs in ALS was conducted, and the convergent validity of the EQ-5D-5L (Herdman et al., 2011), the 5 level version of the EQ-5D-3L, was assessed against the PGI (Ruta et al., 1994), an individualized measure used to identify aspects of ALS that impact patients' HRQL. From this study, results indicated that the majority of GPBMs included only approximately half of the domains identified as important to individuals with ALS. The GPBM with the most coverage was the QWB-SA scale, whereas the HUI3 had the least coverage. Notably, there was no one GPBM that covered all the domains. Therefore, there were domains of impact that were not identified by GPBMs, or that were inaccurately represented and not relevant to our sample.

4.2 Impact of Thesis Results

The results of this thesis highlight the importance of evaluating the psychometric properties of measures when implementing them in a new population (i.e. ALS). It provides an overview of the psychometric properties of GPBMs used in the ALS literature and a reasoning for economic analyses. Furthermore, a platform for further work is provided as the studies conducted found that GPBMs in ALS contain items that are not relevant to individuals, nor do they accurately capture the HRQL of individuals. Therefore, the development of an ALS specific PBM is recommended.

4.2.1 Health-Related Quality of Life in ALS

In this thesis, the PGI (Ruta et al., 1994) was used as a self-report, individualized measure of HRQL to ascertain the areas of patients' lives affected by ALS. The 78 areas identified as important to the HRQL of individuals with ALS were classified into 25 overarching domains; resulting in a total of 25 areas of patients' lives impacted by ALS. These areas were rated by participants as being severely impacted by ALS and included domains such as physical health, recreation and leisure activities, social well-being, self-care, and independence.

It is well known that progression of ALS results in impaired physical function as a result of the degeneration of voluntary motor neurons. Moreover, the impact of decreased physical function is apparent in the domains nominated by our sample. However, other studies have shown that a decrease in physical function, as a result of natural disease progression, is not necessarily important to the overall HRQL of individuals with ALS (Chio, 2004; Goldstein et al., 2002; Neudert et al., 2004; O'Doherty et al., 2010; Robbins et al., 2001; Simmons et al., 2000). The literature has indicated this is due to the impact of patients' mental and social well-being on their ability to cope with these physical changes. It has been theorized and evidenced that this

could potentially be a result of patients' mental fortitude; or an underlying understanding of their diagnosis and the accompanying symptoms and changes that it will bring to their lives. Our study in Chapter Three corroborated the results of the above studies for areas of patients' lives impacted by ALS were found to not only lie in the physical realm but also in the realms of participation, social interactions and mental well-being. These findings therefore detail not only the effects of ALS on patients' HRQL but serve to provide an overview of various domains that measures of HRQL should include in order to accurately assess the HRQL of individuals with ALS. Therefore, the reliability and validity of measures of HRQL in ALS is important to assess and was summarized in Chapter Two and Three.

4.2.2 Quality of Generic Preference-Based Measures Used

Out of the seven leading GPBMs, the EQ-5D-3L (Bansback et al., 2012) and QWB-SA scale (Seiber et al., 2008) were the only two GPBMs that have been used by ALS researchers. In order to accurately assess HRQL, measures should: i) have a defined purpose and definition for the construct of interest (Brazier et al., 2017); ii) be subjective (Fitzpatrick et al., 1998; Mayo, 2015); iii) be based on a conceptual model (Bakas et al., 2012); and, iv) provide a single index score that can be used for comparative purposes and cost-utility analysis (Brazier et al., 2017; Guyatt, 1993). Both measures were developed using general population preferences, whereby the EQ-5D-3L used a TTO method and the QWB-SA scale used a VAS method. Both methods have their advantages and disadvantages; however, both utilize a direct approach to derive general population preferences. HRQL is defined as the construct of interest for each, yet neither are based on any of the three most used conceptual models (Bakas et al., 2012), the Wilson Cleary & Revised models or the ICF. However, the QWB-SA scale was developed using theory from the General Health Policy Model which incorporates mortality and morbidity, as well as preferences

for utilities (Seiber et al., 2008). This might serve to strengthen the measure and could help to explain the greater content coverage observed when the measure was mapped against the domains nominated by participants in Chapter 3. Furthermore, both measures have been deemed reliable and valid in multiple populations (Andresen et al., 1998; Brazier et al., 2004), however in ALS the results are mixed. The responsiveness of both measures has not been assessed; therefore, we cannot draw conclusions on the ability of these measures to detect clinically important change.

As none of the other GPBMs (i.e. SF-6D, AQoL-8D, HUI2, HUI3 and 15D) have had their psychometric properties evaluated, we cannot conclude the overall quality of assessment of GPBMs in this population. However, using an individualized measure to map against all GPBMs (Chapter Three) does allow us to get a sense of their content coverage, the extent to which the majority of GPBMs had about 50% content coverage. One could surmise that based on the results the measures may not fully demonstrate content validity nor validate hypothesized values in their assessment of the remaining psychometric properties.

In Chapter Three, the EQ-5D-5L demonstrated domains that were relevant to participants with 53% content coverage. Yet, it neglected domains pertinent to social well-being and was thus missing a vital domain impacted by ALS. Additionally, as was indicated in Chapter Two, the 3-level version was the most used GPBM in ALS, demonstrating convergent and known-groups validity in ALS. However, the observed floor effects are of particular interest as the measure typically demonstrates ceiling effects in other conditions (Brazier et al., 2004). For example, in MS ceiling effects were reported for the mobility and self-care items in the EQ-5D-3L; and self-care and anxiety/depression items for the EQ-5D-5L (Kuspinar & Mayo, 2014). Similarly, Brazier et al. (2004) also reported ceiling effects when comparing the EQ-5D-3L against the SF-

6D. In the study conducted by Brazier et al. (2004), a greater percentage of individuals reported full health with the EQ-5D-3L compared to the SF-6D. As the authors suggest, this would indicate that the measure is not capable of distinguishing between health states close to full health. The floor effects observed in ALS suggest the EQ-5D-5L is not able to distinguish between more severe health states. As ALS is a fatal neurodegenerative disease, individuals worsen as the disease progresses. The EQ-5D-5L indicates both ceiling and floor effects; however, this may be dependent on which condition the measure is used in. For example, conditions that follow their own natural disease progression, whether it be the relapsing remitting pattern seen in MS or the progressive decline seen in ALS, may show differing limitations from the EQ-5D-5L.

4.2.3 Future Directions

Findings from both studies suggest that GPBMs may underestimate the effects of ALS on HRQL as i) the QWB-SA scale did not demonstrate a high correlation with other GPBMs in ALS when expected to, and ii) the EQ-5D-5L demonstrated floor effects and thus may not be able to distinguish between more severe health states. This is concerning as measures implemented in ALS need to be responsive to changes in disease state. Moreover, decisions regarding treatment evaluation and economic analysis at the population level rely on measures being psychometrically valid for that specific population. This further supports the importance of evaluating the psychometric properties of these measures in the population in which it will be used. Without evaluation, one cannot be sure how accurately scores reflect the construct of interest or which measure is the best to use for treatment decision-making purposes for individuals or groups.

Generic measures are useful to compare across interventions and populations. However, one of the challenges of using generic measures is that they have predetermined domains that tend not to capture all of the domains affected by specific health conditions (Guyatt, 1993; Kuspinar & Mayo, 2013). There may be a loss of relevance in specific contexts and less sensitivity to changes in health status in a specific health condition (Fitzpatrick et al., 1998). As indicated through the work conducted in this thesis, there are domains included in generic measures that are of importance to individuals with ALS. For example, all the domains identified by participants using the PGI were included in at least one GPBM. However, there were also domains not included in GPBMs – in other words none of the measures captured 100% of the areas impacted by ALS. As a result, researchers and health care professionals may use more than one measure in ALS to determine the HRQL of individuals, however this is problematic as the purpose of economic analysis is to provide one measure or index score that can be used to determine the clinical and cost effectiveness of treatments. Furthermore, there needs to be a primary measure as an indicator of outcome. By using more than one measure, there is greater participant burden resulting from disease progression, weakness and fatigue. Therefore, there are two possible solutions to tackle the limitations found with generic measures in ALS: i) adding disease specific "bolt-ons" or dimension extensions to GPBMs or ii) developing a disease specific preference-based measure.

Disease specific bolt-ons are dimensions that can be added to an established instrument in order to overcome perceived inadequacies of the original measure (Longworth et al., 2014). These may improve the validity, precision and responsiveness of these measures in ALS. Bolt-ons have been tested and developed for the EQ-5D for instance, since it is the recommended GPBM for cost-utility analysis. Several domains have been identified as bolt-ons for this

measure, such as the addition of cognition for cancer (Lin et al., 2013) or vision for ocular disease (Longworth et al., 2014). In developing these extensions, the wording and phrasing first needs to be developed and then a valuation exercise is conducted to calculate a scoring algorithm. However, one challenge of the bolt-on approach is that the addition of the new domain may change the values for the original dimensions (Longworth et al., 2014).

Disease specific preference-based measures could also be developed for ALS. As seen in various populations such as stroke (Poissant et al., 2003), asthma (Revicki et al., 1998) and MS (Kuspinar et al., 2016), a disease specific measure is established to include only domains relevant to individuals with the condition. In ALS, areas nominated by individuals using the PGI could be used to develop the domains in the measures and then triangulated against a model like the ICF. In producing a single index score, the measure would therefore provide an accurate evaluation of the clinical effectiveness and cost-effectiveness of different treatment options in this population. Furthermore, disease specific measures tend to be more sensitive and responsive to smaller changes over time than generic measures (Guyatt, 1993).

4.3 Strengths and Limitations

The current studies are not without their strengths and limitations. Both studies were novel and provided information not currently assessed in ALS by summarizing the psychometric properties of GPBMs in ALS and evaluating the content validity of GPBMs in ALS. The review in study one was conducted systematically to strengthen the conclusions that can be drawn from it. Furthermore, the second study was multisite and recruited participants from 3 regions across Canada. The sample was also representative of the ALS population and included a wide range of

ALSFRS-R scores, indicating mild to severe functional impairment, which strengthens the findings from Chapter Three.

While limitations of the individual studies have been described within each of the papers, there are a few limitations to make note of. In addition to a small sample size of included studies in Chapter Two, there was only one study with the primary purpose of psychometric evaluation. Thus, more studies should be conducted with the primary purpose of psychometric evaluation and utilization of GPBMs. A stronger conclusion could then be made regarding the appropriateness of these measures for clinical research and economic evaluation. Future research should therefore focus on implementing GPBMs in ALS and evaluating their psychometric properties.

From the second study, limitations exist in using the ICF as a coding framework as some domains were difficult to code; this lies in the structure of the ICF itself. There was also a large proportion of missing data in completion of the hardcopy surveys. If time and money permitted, completed hardcopy surveys would have been collected in order to enable a stronger validation of the PGI and EQ-5D-5L.

4.4 Conclusion

This thesis contributes to the knowledge base concerning generic preference-based measures and their usage in ALS. It highlights the importance of complete psychometric evaluation and pushes for a greater quality of studies being conducted in ALS. Furthermore, as there were several domains important to be people with ALS not identified by generic preference-based measures, there is a need for the development of an ALS specific preference-based measure that reflects the health concerns of individuals with ALS.

References

Andresen, E. M., Rothenberg, B. M., & Kaplan, R. M. (1998). Performance of a self-administered mailed version of the Quality of Well-Being (QWB-SA) questionnaire among older adults. *Medical Care*, *36*(9), 1349–1360. https://doi.org/10.1097/00005650-199809000-00007

Bakas, T., McLennon, S. M., Carpenter, J. S., Buelow, J. M., Otte, J. L., Hanna, K. M., Ellett, M. L., Hadler, K. A., & Welch, J. L. (2012). Systematic review of health-related quality of life models. *Health and Quality of Life Outcomes*, *10*(1), 134. https://doi.org/10.1186/1477-7525-10-134

Bansback, N., Tsuchiya, A., Brazier, J., & Anis, A. (2012). Canadian Valuation of EQ-5D Health States: Preliminary Value Set and Considerations for Future Valuation Studies. *PLoS ONE*, *7*(2), e31115. https://doi.org/10.1371/journal.pone.0031115

Brazier, J., Ratcliffe, J., Salomon, J. A., & Tsuchiya, A. (2017). *Measuring and valuing health benefits for economic evaluation* (Second edition). Oxford University Press.

Brazier, J., Roberts, J., Tsuchiya, A., & Busschbach, J. (2004). A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics*, *13*(9), 873–884. https://doi.org/10.1002/hec.866

Chio, A. (2004). A cross sectional study on determinants of quality of life in ALS. *Journal of Neurology, Neurosurgery & Psychiatry*, *75*(11), 1597–1601. https://doi.org/10.1136/jnnp.2003.033100

Fitzpatrick, R., Davey, C., Buxton, M. J., & Jones, D. R. (1998). Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment (Winchester, England)*, *2*(14), i–iv, 1–74.

Goldstein, L., Atkins, L., & Leigh, P. (2002). Correlates of Quality of Life in people with motor neuron disease (MND). *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders*, *3*(3), 123–129. https://doi.org/10.1080/146608202760834120

Guyatt, G. H. (1993). Measuring Health-Related Quality of Life. *Annals of Internal Medicine*, *118*(8), 622. https://doi.org/10.7326/0003-4819-118-8-199304150-00009

Herdman, M., Gudex, C., Lloyd, A., Janssen, Mf., Kind, P., Parkin, D., Bonsel, G., & Badia, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*, *20*(10), 1727–1736. https://doi.org/10.1007/s11136-011-9903-x

Kaplan, R., Ganiats, T., Sieber, W., & Anderson, J. (1998). The quality of well-being scale: Critical similarities and differences with SF-36. *International Journal for Quality in Health Care*, *10*(6), 509–520. https://doi.org/10.1093/intqhc/10.6.509

Kuspinar, A., & Mayo, N. E. (2013). Do generic utility measures capture what is important to the quality of life of people with multiple sclerosis? *Health and Quality of Life Outcomes*, *11*(1), 71. https://doi.org/10.1186/1477-7525-11-71

Kuspinar, A., & Mayo, N. E. (2014). A Review of the Psychometric Properties of Generic Utility Measures in Multiple Sclerosis. *PharmacoEconomics*, *32*(8), 759–773. https://doi.org/10.1007/s40273-014-0167-5

Kuspinar, A., Pickard, S., & Mayo, N. E. (2016). Developing a Valuation Function for the Preference-Based Multiple Sclerosis Index: Comparison of Standard Gamble and Rating Scale. *PLOS ONE*, *11*(4), e0151905. https://doi.org/10.1371/journal.pone.0151905

Lin, F.-J., Longworth, L., & Pickard, A. S. (2013). Evaluation of content on EQ-5D as compared to disease-specific utility measures. *Quality of Life Research*, *22*(4), 853–874. https://doi.org/10.1007/s11136-012-0207-6

Longworth, L., Yang, Y., Young, T., Mulhern, B., Hernández Alava, M., Mukuria, C., Rowen, D., Tosh, J., Tsuchiya, A., Evans, P., Devianee Keetharuth, A., & Brazier, J. (2014). Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: A systematic review, statistical modelling and survey. *Health Technology Assessment*, *18*(9). https://doi.org/10.3310/hta18090

Mayo, N. E. (2015). *ISOQOL Dictionary of quality of life and health outcomes measurement*. ISOQOL. https://books.google.ca/books?id=cKjksgEACAAJ

Neudert, C., Wasner, M., & Borasio, G. D. (2004). Individual Quality of Life is not Correlated with Health-Related Quality of Life or Physical Function in Patients with Amyotrophic Lateral Sclerosis. *Journal of Palliative Medicine*, *7*(4), 551–557. https://doi.org/10.1089/jpm.2004.7.551

O'Doherty, L. J., Hickey, A., & Hardiman, O. (2010). Measuring life quality, physical function and psychological well-being in neurological illness. *Amyotrophic Lateral Sclerosis*, *11*(5), 461–468. https://doi.org/10.3109/17482960903552488

Poissant, L., Mayo, N. E., Wood-Dauphinee, S., & Clarke, A. E. (2003). The development and preliminary validation of a Preference-Based Stroke Index (PBSI). *Health and Quality of Life Outcomes*, *1*(1), 43. https://doi.org/10.1186/1477-7525-1-43

Revicki, D. A., Kline Leidy, N., Brennan-Diemer, F., Sorensen, S., & Togias, A. (1998). Integrating Patient Preferences Into Health Outcomes Assessment. *Chest*, *114*(4), 998–1007. https://doi.org/10.1378/chest.114.4.998

Robbins, R. A., Simmons, Z., Bremer, B. A., Walsh, S. M., & Fischer, S. (2001). Quality of life in ALS is maintained as physical function declines. *Neurology*, *56*(4), 442–444. https://doi.org/10.1212/WNL.56.4.442

Ruta, D. A., Garratt, A. M., Leng, M., Russell, I. T., & MacDonald, L. M. (1994). A New Approach to the Measurement of Quality of Life: The Patient-Generated Index. *Medical Care*, *32*(11), 1109–1126. https://doi.org/10.1097/00005650-199411000-00004

Seiber, W. J., Groessl, E. J., David, K. M., Ganiats, T. G., & Kaplan, R. M. (2008). *Quality of Well-Being Self-Administered Scale (QWB-SA)*. San Diego Health Services Research Centre, University of California. https://www.researchgate.net/profile/Kristin_Kistler/publication/252316672_Quality_of_well_being_self-administered_QWB-SA_scale/links/5437d6990cf2590375c55a65/Quality-of-well-being-self-administered-QWB-SA-scale.pdf

Simmons, Z., Bremer, B. A., Robbins, R. A., Walsh, S. M., & Fischer, S. (2000). Quality of life in ALS depends on factors other than strength and physical function. *Neurology*, *55*(3), 388–392. https://doi.org/10.1212/wnl.55.3.388

Smith, P. S., Crossley, B., Greenberg, J., Wilder, C., & Carroll, B. (2000). Agreement among three quality of life measures in patients with ALS. *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders : Official Publication of the World Federation of Neurology, Research Group on Motor Neuron Diseases*, *1*(4), 269–275.