Developing a Cerebral Palsy Preference Based Utility Measure using a Discrete Choice Experiment for cost-utility analyses

Mina Bahrampour BSc, MSc

Centre for Applied health Economics,

School of Medicine,

Griffith University.

A research paper submitted in fulfilment of the requirement for the degree of Doctor of Philosophy at Griffith University, Queensland, Australia.

November 2020

Abstract:

Cerebral palsy (CP) refers to a cluster of developmental disorders of movement and posture. This lifelong disability causes activity restriction and limits the individuals' participation in everyday activities. Further, due to the CP being a lifelong disability, interventions and treatment impose a significant burden on the people with CP, their families and the health care system. Resources are scarce and governments need to know the best way to allocate these resources. A common method to address this requirement is using economic evaluation to provide essential information for resource allocation within the health care sector. Cost-utility analysis (CUA), provides a universal outcome for the economic evaluation of health care interventions that are expressed as a quality-adjusted life-year (QALY). QALYs combines quality and quantity of life into a single index. Quality of life for QALYs is valued by the preference weights given for different health states from the preference based instruments.

Previously published systematic reviews have demonstrated that a limited number of CUAs were conducted for interventions for the CP population. The studies indicated that the current available preference based measures do not perform well in CP population, and there are no condition-specific preference based measures for CP currently available for CUA use in this population. Therefore, the main objective of this research is to develop a preference based measure for CP to be used in CUA. Preference based measures have two components; a health state classification system and a value set for the health states from the classification system.

The first step was to develop the CP-specific classification system. The classification system was derived from a validated and widely used CP quality of life instrument, the Cerebral palsy quality of life (CPQOL). Factor analysis and Rasch analysis were applied to develop the classification system. Rasch analysis and experts opinion were used to evaluate the construct and

content validity of the classification system at the development stage. The classification system consists of six domains: "Social wellbeing and acceptance", "Physical health", "Communication", "Pain and discomfort", "Manual ability" and "Sleep". Each domain is presented by one item and each item has 5 response levels. The new instrument is named the Cerebral Palsy 6 Dimension (CP-6D).

The next step was to generate a scoring (utility) algorithm for the CP-6D. Utility weights are accrued from the preference of people using preference elicitation methods. A Discrete Choice Experiment (DCE) is an ordinal preference elicitation method that asks participants to choose between two or more discrete alternatives. A systematic review was conducted to find the best approach to apply a DCE for valuing health for a multi attribute instrument. The review illustrated that there is not a golden approach to apply a DCE. However, there were some similarities between the studies. For instance, most of the studies used an online survey within the general population. To choose an approach, researchers need to know the features of the value sets that they want to produce before using the value sets in decision making. The systematic review also confirmed that there are no CP-specific instruments that used a DCE to value health.

To develop the utility algorithm for CP-6D, DCE was applied. When using DCEs to value health, they produce utility values on a latent scale; hence, the weights generated from the DCE need to be anchored onto a full health-dead scale to calculate QALYs. To anchor the values, DCEtto method was used. In DCEtto duration is added as an attribute when designing the DCE.

The valuation study was done in two stages; at first, a pilot study was done with a zero prior design, the priors for the final design were estimated from the pilot study data. The main valuation study was conducted as an online survey. The sample was a representative sample of Australian general population in age and sex. A total of 2002 adults aged 18 years and above

completed the survey. The survey included a set of DCEtto tasks, basic social-demographic questions, the CP-6D, and a generic preference based measure (AQoL-4D). DCEtto data were analysed using conditional and mixed logit. All estimated coefficients from the data were in the expected direction and order for all the dimensions. In both methods, all coefficients were negative as expected, except level two of Physical health, which was not statistically significantly different from zero. The Australian utility values of CP-6D ranges from -0.582 to 1.000. The utility algorithm will enable the calculation of utility values for CP-6D in economic evaluations.

The same sample as the DCE study was used to evaluate the construct and criterion validity of the CP-6D, as outlined in the COSMIN checklist. The validity was then evaluated using the AQoL-4D in a general population. There was a moderate correlation between the CP-6D and the AQoL-4D (~0.64), which suggested that the instruments are measuring a similar concept. However, the correlation between domains and items were small, this could be due to different factors. Socio-demographic variable changes had the same effect on both instruments. ANOVA and t-test were used to show the changes in socio-demographic variables; the effect size was medium, and all the results were significant (p-value < 0.001).

This study filled a major gap in the literature in providing health state valuation for CP. A new health state classification system and utility algorithm completes a new preference based Health-Related Quality of Life (HRQoL) measure for CP, however further use and external validation is required. This study has some limitations. Due to the resource constraints and feasibility, the CP-6D was developed to evaluate CP interventions in economic evaluations among the Australian population also like other patient preference settings it can be used for informing patient care and clinical decisions. However, there is a high potential that the instrument to be used in other countries, future research is recommended to estimate the scores in different countries.

Further, the CP-6D should be validated rigorously with people that have CP in a clinical setting. Future studies are recommended to assess whether there is any significant difference between health state utility values for CP-6D derived from the general population and the preferences of people with CP. The CP-6D will facilitate the assessment of CP related interventions using a CUA framework and will assist resource allocation through economic evaluations for CP interventions and treatments, the most common motor disability among children.

THE STATEMENT OF ORIGINALITY

This work has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

…………………………………………..

Mina Bahrampour

November 2020

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

16D	16-dimensional measure of health related quality of life
17D	17-dimensional measure of health related quality of life
AHUM	Adolescent health utility measure
AQL-5D	Asthma quality of life -5 dimensions
AQoL-4D	Assessment of quality of life 4 dimension
AQoL-6D	Assessment of quality of life-6 dimension
BWS	Best-worst scaling
QLU-C10D	Cancer Quality of Life questionnaire
CADTH	Canadian Agency for Drugs and Technologies in Health
CAPI	Computer assisted personal interviews
CP	Cerebral palsy
CPQOL	Cerebral palsy quality of life
CP-6D	Cerebral palsy 6 dimension
CPCHILD	Caregiver priorities and index of life with disabilities
C&CHQ	Care and comfort hypertonicity questionnaire
CHQ	Child health questionnaire
CHU9D	Child health utility-9 dimensions
CTT	Classical test theory
CSPBM	Condition-specific preference based measures
CFCS	Communication Function Classification System
CFA	Confirmatory factor analysis
CA	Conjoint analysis
CM	Conjoint measurement
CUA	Cost utility analysis
CBA	Cost benefit analysis
CEA	Cost effectiveness analysis

CMA	Cost minimisation analysis
DALYs	Disability Adjusted Life Years
DCEs	Discrete choice experiments
DIF	Differential item functioning
EQ-5D	European Quality of Life
EQ-5D-3L	European Quality of Life - 5 Dimensions 3 level descriptive system
EQ-5D-5L	European Quality of Life - 5 Dimensions 5 level descriptive system
EQ-5D-Y	European Quality of Life- youth
EFA	Exploratory factor analysis
GEE	Generalized estimation equation
GDP	Gross domestic product
GPCM	Generalized partial credit model
GRM	Graded response model
GMFCS	Gross motor function classification
HUI(2/3)	Health utilities index (mark 2/3)
HRQoL	Health related quality of life
HTA	Health technology assessment
IRT	Item response theory
KMO	Kaiser–meyer–olkin
MACS	Manual ability classification system
MAUIs	Multi attribute utility instruments
NICE	National Institute for Health and Clinical Excellence
NRM	Nominal response model
OAB-5D	Overactive bladder 5 dimensions
PROM	Patient-reported Outcomes Measures
PBAC	Pharmaceutical Benefits Advisory Committee
PBM	Preference based measures
PCM	Partial credit model
PedsQL	Pediatric Quality of Life Inventory

QALYs	Quality adjusted life years
QoL	Quality of life
QOLCE	Quality of Life for Children with Epilepsy
QLQ-C30	Quality of life questionnaire
QLU-C10D	Quality of Life Utility Measure-Core 10 dimensions
RUT	Random utility theory
RSM	Rating scale model
SQOL-3D	Sexual Quality of Life
SF-6D	Short Form questionnaire-6 Dimensions
SG	Standard gamble
TTO	Time trade-off
VAS	Visual analogue scale
WTP	Willingness to pay
WHO	World health organization
ZBT	Zermelo-bradley-terry

Acknowledgment

Throughout the PhD journey and writing of this dissertation, I have received a great deal of support and assistance.

I would like to express my deepest appreciation to ….

My supervisory team, Professor Paul Scuffham, Associate professor Josh Byrnes and Associate professor Martin Downes for putting their confidence in me to pursue the new and unknown, your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. And Associate professor. Richard Norman for his valuable guidance throughout my studies.

My family, my mum, and dad, for always believing in me and my capacity to overcome any obstacle and to be bold in my curiosity. They never let me give up, even when I was at my lowest. Mahdieh and Soheil, for the non-stop support they gave me, even though they were on the other side of the world.

The Centre of Applied Health Economics (CAHE) and Griffith University, for the input and funding I received towards my PhD and making this journey happen. Ms Gemma Hynard who always guided and assisted me for admin work and made everything easier.

My friends and colleagues, who helped to make this process easier, without whom my PhD journey would have been hard. Especially Ms Connie Gan, for coffee, dinner and walking breaks – you were always there with a word of encouragement or a listening ear.

The Queensland Cerebral Palsy and Rehabilitation Research Centre (QCPRRC) and Professor Roslyn Boyd and Dr Shaneen Leishman for providing some of the data used in this project.

Acknowledgement of published papers

Included in this thesis are papers in *Chapters 2,3,4,5 and 6* which are co-authored with other researchers. My contribution to each co-authored paper is outlined at the front of the relevant chapter. The bibliographic details/status for these papers including <u>all</u> authors, are:

Chapter	Paper
Chapter 2	*Bahrampour M, Byrnes J, Norman R, Scuffham PA, Downes M.* Discrete choice experiments to generate utility values for multi-attribute utility instruments: a systematic review of methods. The European Journal of Health Economics. (21) 983–992 (2020). https://doi.org/10.1007/s10198-020-01189-6
Chapter 3	*Bahrampour M, Downes M, Boyd RN, Scuffham PA, Byrnes J.* Using Rasch and factor analysis to develop a Proxy-Reported health state classification (descriptive) system for Cerebral Palsy. Disability and Rehabilitation. 2020:1-9. https://doi.org/10.1080/09638288.2019.1709565
Chapter 4	*Bahrampour M, Norman R, Byrnes J, Downes M, Scuffham PA.* Developing a cerebral palsy-specific preference based measure for a six-dimensional classification system (CP-6D): protocol for a valuation study. BMJ open. 2019; 9(9):e029325. 10.1136/bmjopen-2019-029325
Chapter 5	*Bahrampour M, Norman R, Byrnes J, Downes M, Scuffham PA.* Utility Values for the CP-6D, a Cerebral Palsy-Specific Multi-Attribute Utility Instrument, Using a Discrete Choice Experiment. Patient (2020). https://doi.org/10.1007/s40271-020-00468-x
Chapter 6	*Bahrampour M, Downes M, Scuffham PA, Byrnes J,* Comparing multi-attribute instruments: CP-6D vs AQoL-4D, Accepted paper - Expert Review of Pharmacoeconomics & Outcomes Research journal

Appropriate acknowledgements of those who contributed to the research but did not qualify as authors are included in each paper.

(Signed) _____ (Date)_____

Mina Bahrampour

(Countersigned) _____ (Date)_____

Supervisor: Prof. Paul Scuffham

1 Introduction

1.1 Background

1.1.1 Cerebral palsy

Cerebral palsy (CP) is one of the most common physical disabilities and movement disorders in children (1). The term cerebral palsy is the combination of the terms cerebral and palsy and these refer to brain and weakness, respectively. This disability encompasses a set of disorders that will cause movement circumscription and permanently affect an individual's ability to move (2); however, the person with CP may have secondary problems such as difficulties with hearing, swallowing, speaking, or vision.

This condition is a life-long disability that occurs due to a non-progressive defect or lesion in the developing brain, during pregnancy or shortly after birth. Its main origins are still unknown. However, some causative factors might be early pregnancy factors (e.g. the mother is exposed to a particular disease), preterm birth, birth asphyxia, multiple pregnancies and post-neonatal brain injuries (3, 4). Even though this disability is permanent, it might change due to other health issues, which are called associative conditions and co-mitigating factors such as autism or asthma. These are not related to CP, but coexist with it (5). Also, children with CP might develop conditions such as sleep disorder or bladder problems. Depending on the management practices and severity of the disease, these may change over time and have an impact (better or worse) on health status.

The disability is not progressive, meaning it does not get worse over time, but as the person gets older, some symptoms may become more or less obvious. It is notable that around 75% to 80% of CP cases are known to happen during the prenatal period and are due to prenatal risk factors (6, 7).

1.1.2 Cerebral palsy categories and severity

Although CP is generally illustrated as a kind of movement dysfunction, the symptoms vary among individuals based on the part of the brain that is injured, and may change over time. Spastic (pyramidal) CP is the most common type in Australia (8) in which the muscles appear tight and the joints are stiff. The most common clinical signs among those who have spastic CP are difficulties controlling movements, poor coordination and balance, and difficulty talking and eating.

There are four types of spastic CP, grouped according to how many limbs are affected (Figure 1.1). Monoplegia occurs when one limb, usually an arm is affected; in hemiplegia one arm and one leg on the same side of the body are affected. Based on the Australian CP Register, these are the most common types of spastic CP. When both legs are affected, it is called diplegia. If both arms and both legs are affected, this is called quadriplegia, and triplegia occurs when both arms and one leg or both legs and one arm are affected (5). Therefore, the level of functioning and quality of life varies between individuals.

Monoplegia   Hemiplegia   Diplegia   Quadriplegia
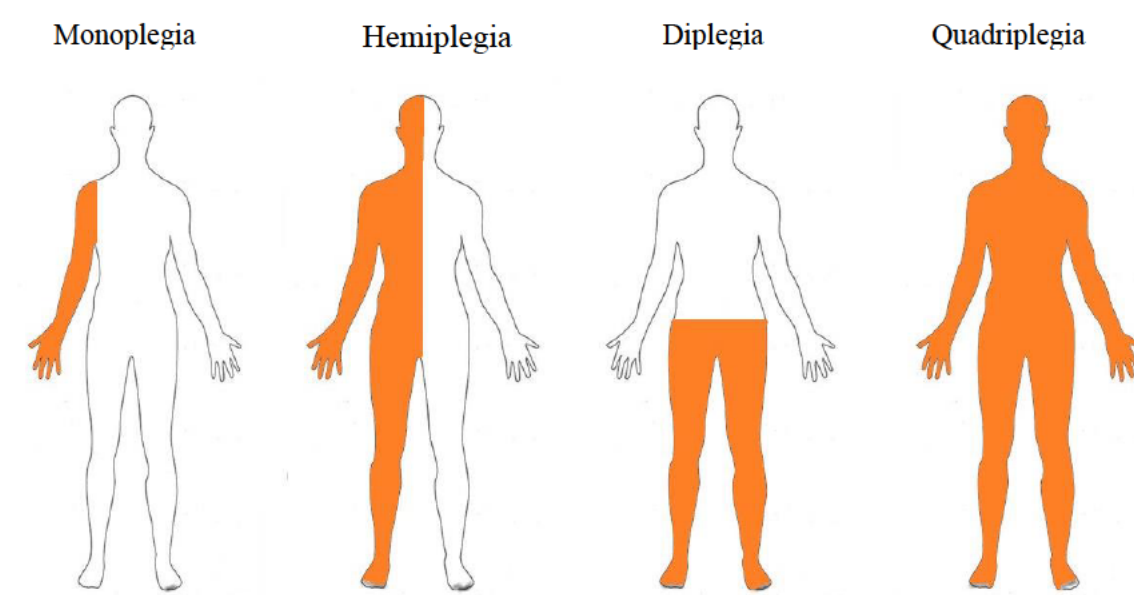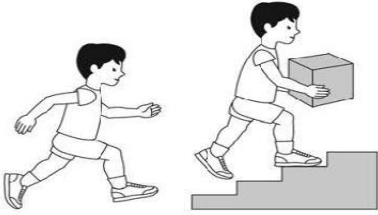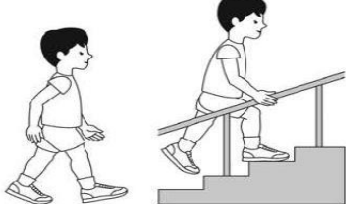
Figure 1.1 Types of cerebral palsy

In addition to movement and posture problems in people with CP, some may have other difficulties such as some level of intellectual disability, impaired vision or hearing, abnormal sensations and perceptions, speech difficulty, seizure disorder and other medical disorders (9).

The life expectancy in CP is different based on the severity of the condition and the quality of treatment that is received. Most children affected by CP live between 30 and 70 years. Children who have mild CP can walk and function more independently and fare better than children with more severe symptoms, such as a low birth weight, limited mobility, and intellectual impairments (10, 11).

CP is a heterogeneous condition, and the functioning of a person with CP is measured through a variety of existing classification systems. The Gross Motor Function Classification (GMFCS) is based on an individual's movement and has five levels. The levels specify the extent of help a person with CP needs and identify whether the person needs a wheelchair or can walk independently (Figure 1.2) (12).

GMFCS E & R Descriptors and Illustrations for Children between their 6th and 12th birthday

GMFCS Level I
Children walk at home, school, outdoors and in the community. They can climb stairs without the use of a railing. Children perform gross motor skills such as running and jumping, but speed, balance and coordination are limited

GMFCS Level II
Children walk in most settings and climb stairs holding onto a railing. They may experience difficulty walking long distances and balancing on uneven terrain, inclines, in crowded areas or confined spaces. Children may walk with physical assistance, a hand-held mobility device or used wheeled mobility over long distances. Children have only minimal ability to perform gross motor skills such as running and jumping.

GMFCS Level III
Children walk using a hand-held mobility device in most indoor settings. They may climb stairs holding onto a railing with supervision or assistance. Children use wheeled mobility when traveling long distances and may self-propel for shorter distances.

GMFCS Level IV
Children use methods of mobility that require physical assistance or powered mobility in most settings. They may walk for short distances at home with physical assistance or use powered mobility or a body support walker when positioned. At school, outdoors and in the community children are transported in a manual wheelchair or use powered mobility.

GMFCS Level V
Children are transported in a manual wheelchair in all settings. Children are limited in their ability to maintain antigravity head and trunk postures and control leg and arm movements.

GMFCS descriptors copyright © Palisano et al. (1997) Dev Med Child Neurol 39:214-23
CanChild: www.canchild.ca

Illustrations copyright © Kerr Graham, Bill Reid and Adrienne Harvey,
The Royal Children's Hospital, Melbourne

Figure 1.2 GMFCS descriptions (source: Palisano et al. (1997) with permission))

The Manual Ability Classification System (MACS) is another categorization tool to describe CP functioning; this represents how much assistance the person needs to use their hands (Table 1.1) (13, 14). Another classificatory tool is the Communication Function Classification System (CFCS); it illustrates how well a person with CP communicates with others (15). CP has extensive effects on all aspects of a person's life and negatively impacts the quality of life (QoL) of those affected (16, 17).

Table 1.1 MACS levels (13)

MACS Level I	Handles objects easily and successfully.
MACS Level II	Handles most objects but with somewhat reduced quality and/or speed of achievement.
MACS Level III	Handles objects with difficulty; needs help to prepare and/or modify activities
MACS Level IV	Handles a limited selection of easily managed objects in adapted situations.
MACS Level V	Does not handle objects and has severely limited ability to perform even simple actions.

There is still no cure for CP. Its current treatment is primarily supportive care which will be long term and challenging. Treatment is dependent on the type and severity of CP. Because of the supportive instead of curative nature of care, the main treatment goal for children with CP is to improve the QoL. Therefore, it is imperative to have a valid and reliable tool to measure QoL in this population, and to be able to measure the efficacy of treatments.

1.1.3 Economic burden of CP

The prevalence of CP in the population is 2.11 per 1,000 live births in high-income countries and around 2 to 2.8 in low and middle income countries (18). Worldwide 17 million people have CP. In Australia, the number of infants born with CP is between 600 and 700 each year (2). A 2007 study reported that for the previous 60 years, CP rates in Australia had remained stable at 2-2.5 cases per 1,000 live births (19). Approximately 34,000 people are living with CP in Australia; and as the population grows, the number of Australians with CP is expected to increase to a predicted 47,600 in 2050 (20).

Access Economics studies showed that CP is one of the top five most costly conditions on a per capita basis (20). The annual financial cost of care for CP in Australia was about AU$1.47 billion (0.14% of GDP) or ~AU$43,431 per person per year in 2007 (20). In 2018, the total financial costs associated with CP were estimated to be AU$3.03 billion, which was

AU\$85,184 per person with CP (21). The cost of the condition has a direct (positive) relationship with its severity, meaning that the more severe the CP (usually based on GMFCS), the more costly the condition becomes. The higher cost might be due to children with severe cases needing more equipment to help them get around in addition to all other expenses related to this condition (22, 23) such as physiotherapy, or the cost of the medicine. A study in 2003 showed that in the US, the average lifetime costs for CP per person were US\$921,000, which includes 80.6% for indirect costs or approximately US\$742,326, 10.2% (US\$93,942) for direct medical costs and 9.2% (US\$84,732) for direct non-medical costs (24, 25). In 2005 the annual Medicaid average cost for a child with CP was US\$43,338 (26).

As the cost for this condition is high, families might need assistance from the government to pay the expenses for CP treatment and care support. Due to the limitation of resources and money, the health care decision-makers need to have dependable information to select the best interventions to increase the QoL and decrease the associated burden to the individual, family and health care system to provide the best value for money. Health technology assessment (HTA) is a multidisciplinary field that uses a range of scientific evidence to assess the quality, efficacy, effectiveness and cost-effectiveness of health care interventions (27). The assessment provides information for health care funding and resource allocation. Health economic evaluation, as a tool of HTA, seeks the best allocation methods of health finances to maximize their impact on health in a resource-scarce setting (28).

1.1.4 Health economic evaluation

One of the main goals of health care delivery is to provide optimal outcomes for patients; where the value is defined as outcomes per dollar spent. Therefore, the ultimate goal is to always search for ways to improve health outcomes with the lowest costs. Economic evaluation is an important tool, and applying economic evaluation helps to find the best and

effective ways to allocate resources (29). Moreover, it leads policy-makers to make effective decisions and inform them about the benefits associated with allocating resources to new health interventions, technologies, health services, practices, and policies (30). Health related quality of life (HRQoL) measures are often used in relation to new health intervention costs in economic evaluation to support decision-making, especially in HTA processes (28).

According to the viewpoint adopted in the analysis, cost and benefit vary. There are different types of economic evaluations: cost-minimization analysis (CMA), cost consequences analysis (CCA), cost-effectiveness analysis (CEA), cost-benefit analysis (CBA), and cost-utility analysis (CUA). Each of these analyses involves a methodical classification and measurement of the costs and consequences of the interventions.

In all economic evaluation methods, the input is always the cost of the intervention, however, the outcomes differ. In CMA, the outcomes of the health interventions are considered broadly equivalent; hence, value for money depends on the intervention that has a lower cost (31).In CCA the outcome or the consequences are health outcomes of a test or s treatment with a suitable alternative. CEA measures the intervention effectiveness using natural units such as improving health status functioning (such as units in blood pressure) or life years (LYs) gained. The difference between CMA and CEA is when using CMA, the outcome is considered equivalent; however, in CEA, the outcome can vary (32). In recent years, CMA has not been considered to be a complete economic evaluation method and it might be inappropriate for many comparisons (31, 32).

The cost and outcomes for CBA are measured in monetary units, where the outcomes can possibly be compared across different health programs. CUA measures the outcome in terms of quality adjusted life years (QALYs), which considers both life years and the quality of those life years (31). The outcome measure of CUA is most often referred to as either

QALYs or disability adjusted life years (DALYs) (33). CEA and CUA, based on the literature, are the predominant methods used in economic evaluation (31). However, the more appropriate method of economic evaluation for health sector decision-makers is CUA (34, 35) because it uses a measure of health outcome that can be compared across all health programs. The most commonly used measure to estimate the outcome (health gain) is QALYs.

Health economic evaluation has adopted the theory of preference utilitarianism. This theory seeks to increase social welfare by maximizing the individuals' utilities (welfarism) or maximise the QALYs (extra-welfarism) (28). The utility can be defined as the pleasure or happiness or satisfaction of preference (36). Welfarism seeks to measure the essential goals of economic activity, however, extra-welfarism claims to illustrate the objectives of the health sector which is maximizing health (28). The QALY maximization principle, along with health economic evaluation principles, follows the utilitarianism principles (37).

1.1.5   Cost-utility analysis

CUA is one of the recommended techniques for evaluation, especially in Australia. Some of the international HTA agencies such as the National Institute for Health and Clinical Excellence (NICE) (38) in the United Kingdom, the Australian Pharmaceutical Benefits Advisory Committee (PBAC) (39) and the Canadian Agency for Drugs and Technologies in Health (CADTH) (40) recommend the use of CUA for government reimbursement decisions. CUA measures resources as monetary values (costs) and outcomes as health gains. CUA typically uses QALY as the summary health outcome measure; the results will be reported as an incremental cost per QALY (41). QALY is a measure to assess the value of health outcomes; it combines the quantity of life (length of life) and QoL in one index (42). QALYs are calculated by multiplying the time spent in each health state by the utility weight associated with that health state (Equation 1) (43).

$$\text{(Eq1)} \quad \text{QALY= utility value (quality of life) x life duration (years of life)}$$

1.1.6 Quality of life

QoL is a multi-dimensional concept; thus, a consensus on its definition has been a significant challenge. The most comprehensive description of QoL is by the World Health Organization (WHO): "individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns." (44). One of the concepts of QoL is HRQoL that focuses on the impact of health status on quality of life. Given QoL's complexity, HRQoL is therefore also a multi-dimensional concept that includes domains related to social functioning, physical, and mental health (45).

HRQoL refers to peoples' or patients' consideration and satisfaction of their current health state and compares it with the health status that they believe is the best or ideal (46). Some researchers may use the terms HRQoL and patient-reported outcomes measures (PROMs) interchangeably (47). An individual's HRQoL also captures the advantages and disadvantages of health related interventions, which are less likely to be identified by standard clinical outcomes. Hence, HRQoL is considered to be a subjective evaluation method, although measuring HRQoL is a vital outcome indicator to evaluate health care interventions and allocate resources to health care.

To assess QoL, instruments can be either generic or condition-specific (48). Generic instruments such as the KIDSCREEN, the Child Health Questionnaire (CHQ), and the Pediatric Quality of Life Inventory (PedsQL) allow the comparison of different patient groups. They can be used in broader samples, and their focus is not on the impact of a specific disease. Generic instruments may lack the sensitivity of disease-specific impacts, and they will not always capture the range of domains in which impairments occur.

Even though generic instruments are intended to represent all relevant aspects of health regardless of the disease area, their validity is questionable for some health conditions (23, 24). Validity is the degree to which an instrument is measuring what it intends to measure. Content validity refers to how well a tool measures the behavior for which it is intended, and convergent validity refers to how thoroughly the new scale is related to other measures that have the same construct. Benefits of treatment and interventions are captured by changes as measured by these instruments; if these instruments lack content validity, they fail to capture elements of importance or interest. If these measures lack discriminant validity, they will be unsuccessful in measuring changes that are important. Therefore, in some conditions based on the disease characteristic, using condition-specific measures may help policy-makers to make more accurate decisions.

Disease or condition-specific instruments, like the Cerebral Palsy Quality Of Life (CPQOL) questionnaire or the Quality of Life for Children with Epilepsy (QOLCE) questionnaire, aim to target a specific disease and its associated symptoms and domains. Hence, condition-specific instruments might be more sensitive to changes in the quality of life for those with that condition and to the impact of that particular disease as they capture the most important domains of QoL. Regardless of the benefits of condition-specific instruments, generic instruments are used more frequently for making comparisons across different health conditions.

Both generic and condition-specific measures can be useful when assessing the efficiency of a treatment or intervention; however, the use of these measures in economic evaluation has its limitations. Both measures require an algorithm associated with them to generate a summary score when comparing interventions and treatments. Some of these algorithms need an indication of the society's preferences for different health states. The

preference can be measured based on the persons' health gain from the intervention or treatment. The health gain can help policy-makers and decision-makers in health care to decide on the best treatment, intervention or health service. One of the most common measures for health gain is QALYs.

1.1.7 Quality adjusted life years (QALYs)

QALYs can combine quality (morbidity) and quantity (mortality) gain within a single summary measure (49). QALYs are calculated by multiplying the duration spent in the health state by utility scores associated with that health state (50). Utility scores are explained below. QALY scores are usually estimated on a scale of 0 to 1, with one being in full health and zero representing dead. QALYs can also have a negative value which means the health state is so bad that it is considered worse than dead (51, 52).

In the concept of QALYs, utility values (in other words, health state preference values) are important. In health economics, utilities are the strength of a person's or society's preference for a health related outcome; in other words, a utility can mean preference (36). Higher utilities are associated with more preferable outcomes (51). Health state utility values or utility weights, to measure QALYs, can be estimated using preference based measures (PBM) or they can be generated through direct elicitation.

1.1.8 Preference based quality of life measures (PBMs)

Preference based measures (PBM), so-called multi-attribute utility instruments (MAUIs) (53), are increasingly being used in utility values estimation. PBMs are used in health related economic evaluations to calculate QALYs. These measures have two components, a system for health state classification (also known as "health state descriptions") and a set of preference weights that are defined by the classification or descriptive system (54).

The classification system typically has a set of domains such as activity, pain, etc. Each domain has some response options which are also referred to as the levels. That is to say, which participants can describe different aspects of their health using a level of each domain (55).

The preference weights produce a single index score for each health state using an algorithm. This algorithm is based on the preference elicitation tasks of a broader population. The algorithm provides a utility value for each health state as described by the descriptive system. The scoring algorithm index generates utilities which usually are scored between zero and one. One refers to full or perfect health, and zero is the health value for death.

PBMs can be either generic or condition-specific. It is considered that PBMs such as the EQ-5D and the Short Form-6 dimensions (SF-6D) are generic. Being generic means that they can be used for a range of health problems and are applicable for most interventions and patient groups. Therefore, the domains are not specific for a certain condition. Importantly, even though some generic instruments may capture some domains indirectly, for instance, EQ-5D might capture hearing loss in terms of change in usual activities, there is a risk that the instrument will fail to capture it completely as this domain is not specifically included in the instrument and yet hearing loss might be an important domain for a condition. The generic instruments are not identical, the domains of health they cover or the levels representing the diversity within the domains may differ, and they might be valued differently using different elicitation techniques (31).

Four PBMs have been used for children with CP to apply a CUA. The Health Utilities Indexes known as HUI-2 and HUI-3, the Assessment of Quality of Life-4 Dimensions (AQoL-4D) and the EQ-5D-3L produced by Euroqol. Of these, HUI-3 was used widely and was suggested to be the most reliable instrument for this condition because of the instrument's psychometric properties (56). However, this index did not cover some crucial domains for CP

such activity limitation. A PBM is required for children with CP, given the unique characteristics of the condition. It was demonstrated that among the five studies that showed the utility values associated with CP in children (57), only one study (58) estimated utility values for the purpose of economic evaluation.

Even though all these instruments have been used, they have limitations in indicating the most important domains affecting children with CP. Hence, when there is no appropriate MAUI for a condition to apply a CUA, there are two options: mapping the scores of a condition-specific on a generic instrument using regression techniques or developing a condition-specific MAUI.

1.1.9 Condition-specific preference based measures

Condition-specific preference based measures (CSPBMs) are usually used when the generic PBMs can be insensitive to the health condition that is being measured. CSPBMs can be developed using an existing HRQoL measure or as a completely new (de novo) measure (59). A drawback of the existing measures is that usually they are complex and large, meaning that they are multi-dimensional and have a significant number of items, which makes these measures impractical for undertaking an elicitation task to develop a preference based scoring algorithm. Even so, the advantage of developing a CSPBM using existing instruments is the estimated utility values can usually be applied for both existing and future data sets of the existing HRQoL measure.

CSPBMs can be derived in six stages. The first four stages are used to develop a health state classification system. These stages are establishing the number of dimensions, selecting items for each dimension, reducing the items levels and validating the survey. This system reduces the number of health states to produce new values without losing much information (60). The last two stages are for the valuation of the health states as described by the health

state classification system which are modelling health state values and Developing algorithm to obtain utility values.

1.1.10 Cerebral Palsy Quality Of Life instrument (CPQOL)

The CPQOL is a disease-specific instrument, designed to measure HRQoL in children and adolescents with CP. It has been reported as one of the most reliable measures among disease- or condition-specific instruments for children with CP (61). This instrument measures domains related to this disease, such as social wellbeing and acceptance, participation and physical health, feelings about functioning, emotional wellbeing and self-esteem, pain and the impact of the disability on wellbeing, through questions with a scale of 1 (very unhappy) to 9 (very happy) (62).

At first, there were two versions of this questionnaire, one for children and the other for primary caregivers. Now in addition to these, there is the more recently developed questionnaire designed for teens (13-18 years old). CPQOL has a self-reported version and a parent/proxy version (63). In the parent/proxy version two extra domains, access to services and family health, are included. The parent-proxy version (for parents of children aged 4-12 years), contains 65 items. The child self-report version (for children aged 9-12 years) includes 53 items (62). The starting age for CPQOL was chosen as 4 years of age because it ensured that the child has a clear diagnosis of CP. The teen version was developed as the new issues for people with CP might occur during adolescence and new dimension were needed (64).

1.1.11 Utility values for children with cerebral palsy

Measuring health state values in children is a developing field of research (65). There are many well-established techniques for measuring health utilities in adults, but a lot of these methods are not justified to be used for children. It is understandable that the methods used for adults may not be suitable for children and the instruments used might have dimensions that

are not appropriate for children. As the instrument development stages are different among age groups it will be hard to adjust the MAUI dimensions that can be suitable for all the groups (66). For instance, the dimensions important to a 15-year-old might be very different to that of an 8-year-old. Children understand health different from adults; for instance the children may pay more attention to well-being while adults focus on absence of chronic illness (67).

Nine universally accepted generic MAUIs were identified that have been applied in the pediatric population, with most of them being adjusted from the previously existing adult instrument (68). These instruments were the Quality of Wellbeing Scale (QWB), HUI-2, the HUI-3, the 16-dimensional measure of HRQoL (16D), the 17-dimensional measure of HRQoL (17D), the Assessment of Quality of Life-6 Dimensions Adolescent (AQoL-6D Adolescent), the Child Health Utility 9D (CHU-9D), the EuroQoL-5 Dimensions Youth (EQ-5D-Y) and the Adolescent Health Utility Measure (AHUM).

A background analysis of the valuation studies demonstrated that there were no studies that used a CP-specific measure to value health states for CP interventions (57, 69), and the inappropriateness of a generic instrument was mentioned (56). Therefore, a CP-specific measure is required. The first step to a PBM is developing a CP-specific classification system

1.1.12 Developing a health state classification system

A health state classification system, also known as a health state descriptive system, is the first stage toward a complete PBM (70). The classification is constructed from a number of domains and their levels. For this purpose, individuals are asked to respond to the instrument questions, where they are asked about what level they are at for each domain. When the person replies to these questions, they are categorized into what is called a 'health state' (55). Health state classification systems have been developed to help reduce the large number of items from existing QoL instruments.

The first step in the classifying system is to establish the number of domains or factors by applying a factor analysis on items. Factor analysis is a multivariate statistical technique which is commonly used in different fields of studies such as health, psychology, education and applied health economics, to examine the structure or relationship between variables when there are complicated associations. It helps to collapse or reduce a large number of variables to fewer explainable underlying ones (60).

The second step of developing a classification system is selecting items for each dimension. To do so, item response theory (IRT) and Rasch analysis can be used (60). Rasch analysis produces a set of items, which are selected from a survey, to design and use a scale in which the item difficulties and person measures are determined.

The third step is reducing the items and collapsing the levels (if needed), based on the results of the Rasch analysis. This selection can be based on the correlation between items and using clinical and patient opinions. Typically, one or two items will be selected per domain.

The fourth step is the survey validation, and this step can be applied differently based on the data: it can be done by repeating the analysis in the preceding stages on an alternative sample from the same dataset, a sample at a different time point from the same dataset, or use a substitute data set. The validation can also occur in new datasets and will be useful in comparing the performance of other instruments. One of the checklists to validate the instrument is COSMIN (COnsensus-based Standards for the selection of health status Measurement INstruments). COSMIN is a checklist that was developed to evaluate the methodological quality of studies on measurement properties of PROMs. The items that COSMIN checklist discusses are internal consistency, content validity, hypotheses (aspect of construct validity), criterion validity, and responsiveness (71).

1.1.13 Health state valuation

The last two steps of deriving CSPBMs are the valuation, based on survey and modelling. The first step (fifth step of CSPBM development) is defining health states of interest, and then a sample of health states will be chosen from the classification system, often using a statistical approach. In the second step (sixth step of CSPBM development), the health states are valued on a scale where the best (full) health state is anchored as one and "dead" is anchored as zero. In some studies in the scoring system, a negative utility value is also possible; this value shows a health state so poor that it is even worse than death. There are several methods to derive utility values for a MAUI. These methods can employ cardinal preferences such as the standard gamble (SG), time trade-off (TTO), and visual analogue scale (VAS). Or they can be ordinal methods such as the discrete choice experiment (DCE) and a ranking exercise such as best worst scaling (BWS).

1.1.13.1 Standard gamble, time trade-off, and visual analogue scale

The SG and TTO are models to measure cardinal preferences, especially in health care. The SG was first presented by Neumann and Morgenstern (72) and was used to value the SF-6D. The TTO was introduced by Torrance et al (73) and was developed to be less complicated compared to SG. The TTO is the most used method to derive preferences for many instruments. For example, the TTO method was used to develop the scoring algorithm for the EQ-5D, the most widely used MAUI (74).

The SG and TTO are choice-based approaches in which the SG asks about choosing between the health states until a point where the individual is indifferent about the choice between the probabilities of immediate death or living in a specific health state (75). The TTO asks individuals, about how many years of living in full health followed by an immediate death they would be willing to trade, compared to living in a health state for a definite number of

years. It determines that the individual is willing to accept living in a full health state for fewer years relative to being sick for more years.

In the VAS, the participants are asked to value a health state on a scale of zero to 100, where the top of the scale indicates the 'best health' and the bottom indicates 'the worst health.' Even though the VAS is simpler and much easier to understand compared to the SG and TTO, it does not involve the time spent in each health state and does not explicitly involve choices or trade-offs (76). It has been argued that risk aversion and time preference can bias an individual's response when using both the SG and TTO (77, 78).

1.1.13.2 Discrete choice experiments and best worst scaling

DCEs and BWS are usually accomplished without an interviewer and can be applied using an online survey, which expedites the data collection process, therefore the use of ordinal methods has increased in recent valuation studies (79). DCE is an ordinal method that is based on a theory of choice behaviour originally proposed by Thurstone (80) called the Random utility theory (RUT). McFadden (81) then extended the theory from pair-wise comparisons to multiple comparisons.

The RUT can consider inter-linked behaviours and is different from the more traditional conjoint analysis (CA). CA is based on conjoint measurement (CM) and is not a behavioural theory, whereas the RUT explains the choice behaviour of humans, not numbers (82, 83). The DCE model asks participants to choose their preferred health state among alternative hypothetical scenarios (84). The results of DCEs can help to produce values for health states derived from logistic models or models with logit as a link function. If the duration is also included, this model can anchor the latent values onto a full health-dead utility scale. However, if the duration is not added the anchoring for DCEs can also be performed as a separate task. DCE techniques within the context of developing a utility measure are described as follows.

There are several key stages in performing a DCE. First, identify relevant domains of the health outcome measure and assign levels to these domains. Once the DCEs are completed, the data are analyzed using statistical methods to establish the utility weights. These weights are generated using an algorithm that is based on values obtained from a sample (usually the general population); in addition, the weights can be used in an economic evaluation. Different methods are used in DCEs to value health states. These differences can be in terms of design, anchoring, or analysis methods. A background analysis of the best methods to apply DCEs to value health states in MAUIs revealed that a comprehensive systematic review of the subject has not yet been undertaken.

1.1.14 Who can value the health states

There is an ongoing debate when valuing health states about who should value the health states for economic evaluation. One view by researchers is that the people who are familiar with the condition (whether through self-reporting or proxy-reporting) should value the health states. It has been stated that measurement using the patient view is a better option compared to hypothetical preferences, as it is assumed that it could control its practical implications (85). Another argument is to use the values derived from the general population (public) who are taxpayers. As health care services are usually tax-funded, the priority should go to the taxpayers. Reimbursement agencies, for example, the PBAC (39), the NICE (38), and the CADTH (40) prefer the utility weights derived from the general population.

However, where the national guidelines are not available, it is usually up to the researcher to verify whose values have been included in the valuation study. Some studies show that patients give higher utilities to a particular health state than the general population (86-88), while some studies indicate that, for instance, patients with dementia and their proxies gave systematically lower values to some health states than individuals from the general population

(89). The question about who should value health remains one of the most debated areas in health economics (90).

Another argument is whether the health state should be valued by the patient themselves or by their proxy, especially when assessing children's health states. It is assumed that children with CP can report their HRQoL (but this may not be precise). This, however, becomes more challenging in children with more severe cases or comorbidities. Because the validity of the child-reported HRQoL is questionable, usually a proxy report is used for this condition. A review illustrates that in MAUIs that were used for children and adolescents with CP, the majority (88%) were completed by a proxy (e.g. parent or guardian) (56).

Children could lack the ability to report their health states as they might have difficulty understanding the HRQoL questions. In these cases, some researchers are willing to adopt a proxy to complete the measurement task. The proxy can be the child's parent, guardian, primary caregiver or even the child's physician/doctor. Studies have shown that parents are the best proxies because they know the child's situation better than anybody else (59), as doctors do not interact with the child outside of the health care facility or clinic. Previous research illustrates that there is a slight issue with using a proxy for children with disabilities, as the proxy HRQoL data scores are often systematically lower than the self-reports by children (91-93).

1.2   Objectives of this study

This overall research aim is to develop an algorithm and utility values for people with CP that can be used in economic evaluation. To accomplish this aim, the main objectives of this study are as follows:

1) To conduct a comprehensive systematic literature review to identify the multi-attribute utility instruments that use discrete choice experiments as a method to derive preference based values.

2) To develop a disease-specific health state classification (descriptive) system for children with CP, based on the CPQOL, a CP-specific QoL instrument.

3) To generate preference based utility weights for the health states defined by the health state classification system derived from the CPQOL instrument, using DCE.

4) To check the validity of the CP-6D against the AQoL-4D.

1.3   Structure of this thesis

Chapter One presents an introduction and background to this research. In this chapter, the overview of cerebral palsy and its burden are presented. Moreover, this chapter includes a narrative review of the concept of the HRQoL, condition-specific preference based measures, health state classification systems, and economic evaluation. Furthermore, at the end of the chapter, the objectives are illustrated.

Chapter Two provides the findings of a systematic literature review of methods when using discrete choice experiments to value health described by multi-attribute utility instruments. This chapter is prepared in the form of a manuscript and has been published in the academic journal 'European Journal of Health Economics'.

Chapter Three illustrates the development of the CP classification system from CPQOL; this chapter has been published as an original article in the 'Disability and Rehabilitation' journal. Chapter Four introduces the chosen methodology for the valuation study and has been written in a protocol format which has been published in the 'BMJ Open'.

Chapter Five presents the CP-specific instrument (CP-6D) valuation study using DCE to estimate utility values using the general population values. This chapter has been published as an original article in the 'The Patient: Patient-Centered Outcomes Research' journal. Chapter Six shows how the CP-6D was validated using another MAUI, AQoL-4D.

In Chapter Seven, the findings of the studies are discussed, and conclusions from this thesis are presented. The limitations of the research and opportunities for future research are also discussed in this last chapter.

The final part of this document, the Appendix, presents supportive materials. The Appendix provides several summary tables referred to in the chapters, such as the systematic search terms and a summary of literature characteristics from the systematic review, tables that could not fit in the articles published, etc.

This thesis is prepared in accordance with the relevant Griffith University policies (griffith.edu.au/hdr/thesis preparation). Figure 1.3 shows the schematic overview of the structure and content of the thesis.

	Chapter 1: Introduction -Background & research objectives

Objective 1	**Chapter 2: Published Manuscript** - Discrete choice experiments to generate utility values for multi-attribute utility instruments: a systematic review of methods

Objective 2	**Chapter 3: Published Manuscript** - Using Rasch and Factor Analysis to Develop a Proxy-Reported Health State Classification (Descriptive) System for Cerebral Palsy

Objective 3	**Chapter 4: Published Manuscript-** Developing a cerebral palsy-specific preference-based measure for a six-dimensional classification system (CP-6D): protocol for a valuation study	**Chapter 5: Published Manuscript-** Utility Values for the CP-6D a Cerebral Palsy Specific Multi Attribute Utility Instrument using Discrete Choice Experiment

Objective 4	**Chapter 6: Accepted Manuscript-** Comparing multi-attribute instruments: CP-6D vs AQol-4D

|	
	Chapter 7: General discussion, conclusion, and recommendations

Figure 1.3 Schematic overview of the structure and content of the thesis

2   Chapter 2: Discrete choice experiments to generate utility values for multi-attribute utility instruments: a systematic review of methods

STATEMENT OF CONTRIBUTION TO CO-AUTHORED PUBLISHED PAPER

This chapter includes a co-authored paper. The bibliographic details of the co-authored paper, including all authors, are:

Bahrampour M, Byrnes J, Norman R, Scuffham PA, Downes M. Discrete choice experiments to generate utility values for multi-attribute utility instruments: a systematic review of methods. The European Journal of Health Economics. (21) 983–992 (2020).

https://doi.org/10.1007/s10198-020-01189-6

Copyright statement: Copyright of this manuscript belongs to The European Journal of Health Economics (EJHE). Authors permitted to use for scholarly purposes, personal use and internal institutional use. Also, authors can share their preprint anywhere at any time. Further details are available: https://www.springerpub.com/journal-article-sharing-policies

I contributed to the original idea of the study, literature search, quality assessment of the included studies, data analysis and drafting the article.

(Signed) _____ (Date)_____

Mina Bahrampour

(Countersigned) _____ (Date)_____

Corresponding author of paper: Mina Bahrampour

(Countersigned) _____ (Date)_____

Supervisor: Prof. Paul Scuffham

ORIGINAL PAPER

Discrete choice experiments to generate utility values for multi-attribute utility instruments: a systematic review of methods

Mina Bahrampour[1] · Joshua Byrnes[1] · Richard Norman[2] · Paul A. Scuffham[1] · Martin Downes[1]

Abstract

Objectives In recent years, discrete choice experiments (DCEs) have become frequently used to generate utility values, but there are a diverse range of approaches to do this. The primary focus of this systematic review is to summarise the methods used for the design and analysis of DCEs when estimating utility values in both generic and condition-specific preference-based measures.

Methods Published literature using DCEs to estimate utility values from preference-based instruments were identified from MEDLINE, Embase, Cochrane Library and CINAHL using PRISMA guidelines. To assess the different DCE methods, standardised information was extracted from the articles including the DCE design method, the number of choice sets, the number of DCE pairs per person, randomisation of questions, analysis method, logical consistency tests and techniques for anchoring utilities. The CREATE checklist was used to assess the quality of the studies.

Results A total of 38 studies with samples from the general population, students and patients were included. Values for health states described using generic multi attribute instruments (MAUIs) (especially the EQ-5D) were the most commonly explored using DCEs. The studies showed considerable methodology and design diversity (number of alternatives, attributes, sample size, choice task presentation and analysis). Despite these differences, the quality of articles reporting the methods used for the DCE was generally high.

Conclusion DCEs are an important approach to measure utility values for both generic and condition-specific instruments. However, a gold standard method cannot yet be recommended.

Keywords Discrete choice experiment · Conjoint analysis · Health state valuation · Preference-based measures · Utility · Systematic review

JEL Classification I10

✉ Mina Bahrampour
mina.bahrampour@griffithuni.edu.au

1 Centre for Applied Health Economics, School of Medicine and Menzies Health Institute Queensland, Griffith University, Nathan, QLD, Australia

2 School of Public Health, Curtin University, Perth, WA, Australia

Introduction

Resources in health-care are scarce [1]. Finding an efficient way to assign these limited resources has been one of the objectives of economic evaluation. Economic evaluations, which are designed to determine the most efficient and cost-effective interventions [2], generate evidence for prioritising the best set of interventions to maximise total health. To do this, a number of different methods are available. These methods compare the interventions and programs through two dimensions, cost and consequences. The consequences can be either monetary or non-monetary values [3].

One popular method used in economic evaluation is cost–utility analysis (CUA) [4] in which the main outcome is defined with respect to quality adjusted life years (QALYs)

⚫ Springer

Chapter 2: Discrete choice experiments to generate utility values for multi-attribute utility instruments: a systematic review of methods

2.1 Abstract

Objectives: In recent years, discrete choice experiments (DCEs) have become frequently used to generate utility values, but there are diverse range of approaches to do this. The primary focus of this systematic review is to summarise the methods used for the design and analysis of DCEs when estimating utility values in both generic and condition-specific preference based measures.

Methods: Published literature using DCEs to estimate utility values from preference based instruments were identified from MEDLINE, Embase, Cochrane Library and CINAHL using PRISMA guidelines. To assess the different DCE methods, standardised information was extracted from the articles including the DCE design method, the number of choice sets, the number of DCE pairs per person, randomisation of questions, analysis method, logical consistency tests, and techniques for anchoring utilities. The CREATE checklist was used to assess the quality of the studies.

Results: A total of 38 studies with samples from the general population, students and patients were included. Values for health states described using generic multi-attribute instruments (MAUIs) (especially the EQ-5D) were the most commonly explored using DCEs. The studies showed considerable methodology and design diversity (number of alternatives, attributes, sample size, choice tasks presentation, and analysis). Despite these differences, the quality of articles reporting the methods used for the DCE was generally high.

Conclusion: DCEs are an important approach to measure utility values for both generic and condition-specific instruments. However, a gold standard method cannot yet be recommended.

2.2 Introduction

Resources in health-care are scarce (94). Finding an efficient way to assign these limited resources has been one of the objectives of economic evaluation. Economic evaluations, which are designed to determine the most efficient and cost-effective interventions (95), generate evidence for prioritising the best set of interventions to maximise total health. To do this, a number of different methods are available. These methods compare the interventions and programs through two dimensions, cost and consequences. The consequences can be either monetary or non-monetary values (31).

One popular method used in economic evaluation is cost-utility analysis (CUA) (96) in which the main outcome is defined with respect to quality adjusted life years (QALYs) (97). The estimation of QALYs requires utility values for defined health states (98). Usually, these values, which are also known as preference scores, utilities or weights (99), are anchored on a scale where one is representative of full health and zero represents death (100). To ensure comparability between different areas of health, health states are often defined using a multi-attribute utility instrument (MAUI). MAUIs can be either generic or disease-specific (101) and the decision on what kind of MAUIs to use to value any particular health state is still debatable.

MAUIs describe health states based on dimensions of quality of life and are associated with a scoring algorithm to estimate the relative value of that health state. These values are then multiplied by the duration in each health state to calculate QALYs (29). Frequently used preference based methods can be classified into either cardinal methods or more recently, ordinal methods. Cardinal methods for utility estimation assume preference can be expressed quantitatively; however, ordinal methods, such as ranking models (102) and Discrete Choice

Experiments (DCEs) (103-105), do not rely on an individual providing a score for a particular health state; rather they are based on rankings instead (76). In SG, the participants are asked the risk they are going to take for achieving better health and in TTO participants are asked about the amount of time they are willing to trade off for better health. There are some concerns with using SG and TTO preference based methods (106, 107). The tasks in SGs and TTOs might be more difficult to administer in comparison with DCE (108). Secondly, valuing health states worse than dead can be problematic with these techniques (109). These problems have led to the use of ordinal methods such as DCEs (110); however, these methods have their own limitations (79, 111, 112). A DCE is a preference based technique that asks individuals to choose between health states over a set of different hypothetical scenarios in order to elicit their preferred health state and the relative weights for various attributes embodied within health states. DCEs for MAUIs are similar to TTO and SG, in the case that all of the models are presented by their attributes to describe health states. However, DCEs are generally easier to administer and they have greater reliability due to reduced measurement error (106, 108, 113). The derived values from DCEs are based on the preferred health state or health profile elicited by the choices made by respondents (114).

Even though DCEs are predominantly applied to elicit the relative strength of preferences of different attributes within programs or between programs, the use of DCE to value health states for estimating utility weights is increasing (115). There are key stages to perform a DCE for a valuation study. Before starting to design the DCE, the first step is identifying relevant dimensions of the health outcome measure and assigning levels to these dimensions. The dimensions and their levels will be described when using a classification (descriptive) system. To design the DCE the next step is applying statistical design theory to draw an independent sample of health states from the full set of health states, for which

preferences will be elicited. Then the choice sets will be presented to respondents, and they will be asked to express their preferences by choosing within these choice sets.

There are many different methodological and design considerations when using a DCE to value utilities. The validity of these methods is uncertain. Indeed, the methods to derive DCEs are still evolving, and hence it is important to understand how researchers are developing and using DCEs to help define and inform best practice for future studies. This paper reports a systematic review based on methods to determine how DCEs are conducted to estimate utility values for MAUIs. The main aim of the review is to summarise the methods used to design and analyse DCEs when generating utility weights in both generic and condition-specific preference based measures.

2.3 Methods

This systematic review is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols (PRISMA-P) guidelines (116). The protocol for this review was registered in the international database of prospectively registered systematic reviews in health and social care (PROSPERO), Centre for Reviews and Dissemination, University of York (CRD42017075603).

Search strategy

In order to identify the relevant published literature, multiple electronic databases were included in the search strategy: Medline, EMBASE, Cochrane Library and CINAHL. The selected studies were restricted to English language articles published in peer-reviewed journals. References of the included articles were cross-checked to identify other relevant publications. There were no limits on time of publication.

A preliminary scoping search was conducted to identify terminology for the search terms and the type of studies that are likely to be available and the research team discussed and approved the list of key search terms. The main search terms were related to commonly used DCE terminology, quality of life terminology and different multi-attribute instruments; a full description of the search strategy can be found in the supplementary file. Inclusion and exclusion criteria:

Studies had to meet the following criteria to be included:

- Use of a DCE for preference elicitation; and

- Use of an existing MAUI

Studies were excluded if:

- Health related quality of life (HRQoL) was valued using only non-DCE techniques such as TTO or SG;

- A DCE was used, but did not report the results of the DCE design in their study; or

- Results were based on simulated data

More generally, methodological articles with no data, systematic reviews, working papers, protocols, editorials or letters, and abstracts were excluded from the identified articles.

Study selection and quality assessment

MB reviewed article titles and excluded those that clearly didn't match the inclusion criteria. MB reviewed the abstracts that were remaining to assess if they met the inclusion/exclusion criteria. After exclusion by title and abstract, MB reviewed the full text of the remaining studies and decided on the final included article. SK assessed all the articles at

the abstract stage to ensure consistent selection/rejection. If there were any differences in agreement adjudication was undertaken by MD. Assessment of the quality of the articles in this review has been done using the CREATE checklist (117). This checklist was designed to assess the quality of valuation studies of multi-attribute utility-based instruments. The CREATE checklist includes 21 items in seven categories: descriptive system; health states valued; sampling; preference data collection; study sample; modelling; and scoring algorithm. For this study, each item in the checklist was scored either yes or no; however, some items were not applicable in some studies, such as studies that compared models but did not report the algorithms. For each study, the percentage score of reported items was calculated (yes was scored as 1 and no as 0).

Data extraction

Data were extracted from eligible articles using a predefined data extraction sheet (supplementary file). This file included the general information of the studies such as study country, study year; DCE design methods such as the number of choice sets, number of DCE pairs, randomisation of questions, logical consistency tests, and techniques for anchoring utilities; it also consisted of questions regarding other preference eliciting methods conducted, analysis software and statistical models. These categories were chosen as they have been used in previous reviews and have been included in checklists to develop DCEs (118, 119).

2.4   Results

Selection process

After initial searches were completed, a total of 3,162 studies were identified. After reading of titles and abstracts, 141 articles were identified as potentially meeting the inclusion criteria. In the full article review stage, 38 articles were included in the final evidence synthesis (Figure 2.1).

Study descriptive data

From the literature search, a total of 38 articles were used in the systematic review (Figure 2.1).
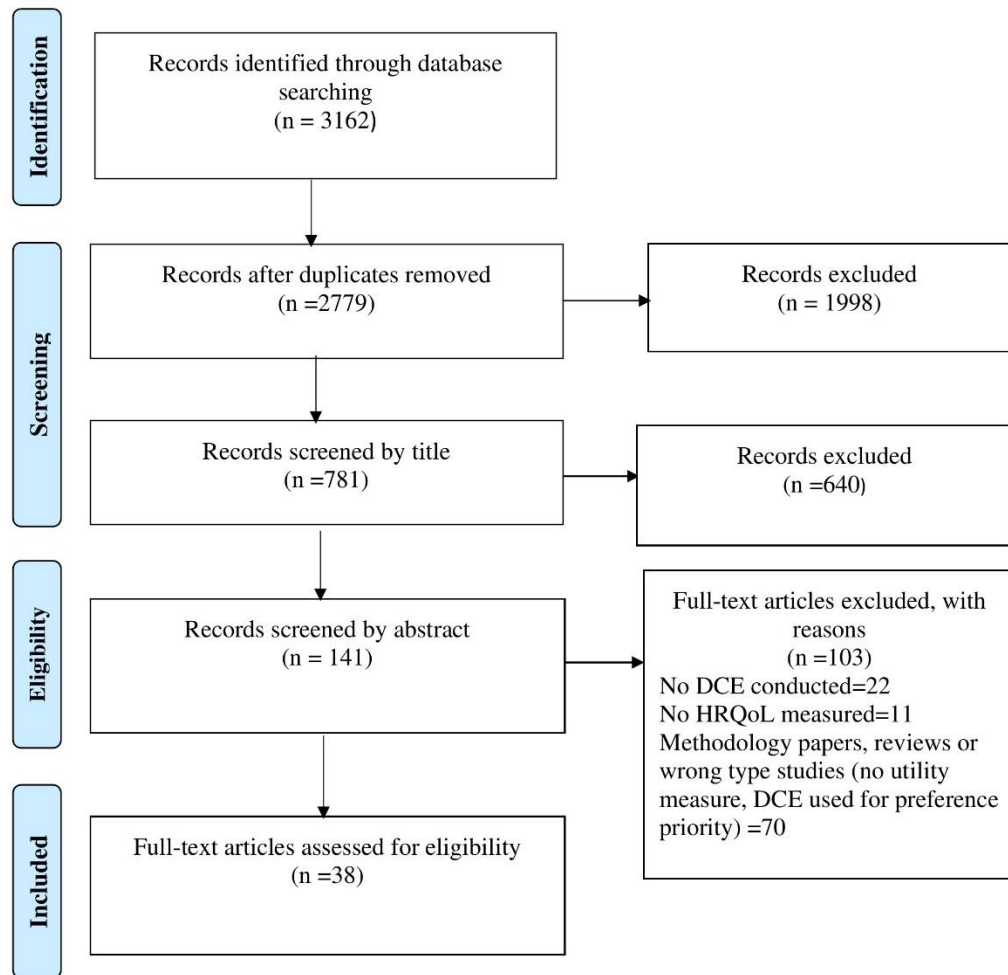
Figure 2.1 PARISMA flow diagram summarizing the study selection process

The EQ-5D, a generic MAUI, was the most commonly used instrument (n=24) for valuation of health states through a DCE, of which 17 studies used the more recent EQ-5D-5L, seven studies used the EQ-5D-3L, and one study used both (120). Other generic MAUIs were the SF-6D (n=2) and the CHU-9D (n=1). Condition-specific MAUIs included cancer-specific instruments (the QLU-C10D (112, 121-123), derived from the EORTC QLQ-C30, (n=4),

Asthma Quality of life (AQL-5D) (104) (n=1), an influenza-related health-related quality of life measure (n=1), the overactive bladder questionnaire (OAB-5D) (78, 104) (n=1), sexual quality of life (SQOL-3D) (n=1), a Glaucoma utility index (n=1), a Diabetes Health Profile-Five Dimension (n=1), and the ICE-CAP Supportive Care Measure(124) (n=1) (Table 2.1).

The majority of studies recruited participants from a single country: the UK (n=12), Australia (n=7), the Netherlands (n=4), Germany (n=2), Spain (n=2), the USA (n=2), Canada (n=1), Indonesia (n=1) and Japan (n=1). Six studies were carried out in more than one country: Xie et al (125) included study participants from the UK, Canada, Spain, the Netherlands, China, Japan, Korea, and Thailand; two studies (122, 123) recruited participants in France and Germany; Krabbe et al studies (99, 126) recruited participants in Canada, England, The Netherlands and the USA, and Pullenayegum et al (74) recruited participants in Canada and the UK. Thirty-three of the studies sampled from the general population, stratified by age and gender (supplementary file Table A 2.1.1). Scalone et al (127) sampled university students in the Netherlands. Stolk et al (106) elicited values from both a general population sample and students. Ratcliffe et al (128) used adolescents as the study sample, as their questionnaire was designed for young adults. Burr et al (114), which used a disease-specific MAUI, used patients to elicit values. Sample sizes ranged from 102 to 8,222 respondents (Supplementary file Table A 2.1.2).

Table2.1 Background information of the studies

Instrument		N=38
MAUIs:	**Complete name**	
EQ-5D-5L	European Quality of Life 5 Dimensions 5level	18
EQ-5D-3L	European Quality of Life 5 Dimensions 3 level	8
QLQ-C30 (QLU-C10D)	Quality of life questionnaire	4
SF-6D	Short-Form questionnaire-6 Dimensions	2
CHU-9D	Child Health Utility-9 Dimensions	1
AQL-5D	Asthma Quality of Life -5 Dimensions	1
OAB-5D + AQL-5D	over active bladder quality of life-5 dimensions	1
SQOL-3D	Sexual Quality of Life 3 dimension	1
ICECAP SCM	ICEpop CAPability measure Supportive Care Measure	1
Diabetes Health Profile-Five Dimension	Diabetes Health Profile	1
Influenza-related health-related quality	Influenza-related health-related quality	1
Number of choices		
	<8	8
	8-12	15
	13-16	12
	16<	3

*One extra MAUI as one study used both EQ-5D-3L and EQ-5D-5L

Administration

Commercial providers administered the majority of the questionnaires (Table A 2.1.1). 30 studies used an online panel to gather data, four studies used interviews and follow up postal questionnaires and two studies used computer assisted personal interviews (CAPI), two studies did not mention their data gathering method.

Modelling DCE data

A number of the included studies developed a task beyond a simple comparison of two or more health states. For example, nine studies added the "duration" of being in a given health state to their DCE questionnaire (DCEtto); 12 studies included a separate TTO experiment of which one also included a willingness to pay (WTP) task (129), two articles included a ranking task (107, 115); and two studies (106, 130) included a visual analogue scale (VAS) task in addition to a TTO and DCE. Two studies (126, 131) included a VAS and seven studies also included a best-worst-scaling (BWS) task in addition to a DCE, the BWS type has not been reported. (Supplementary file Table A 2.1.2). Seven studies used dead in their survey as a third option in the choice set.

34

Attributes and levels

The number of possible health states covered by the instruments ranged from 64 to 2 million. The number of choice sets included in the studies ranged from 24 to 1,620. The DCE choice tasks per respondent varied between 6 and 32 choices (Supplementary file Table A 2.1.1).

Design

The way that DCE designs were reported in the studies varied; the reported methods were full factorial (n=1), fractional factorial (n=3), efficient design (d-efficient and two-way interaction; n=11), Bayesian methods (n=12), d-optimal (n=5) and orthogonal methods (n=3). In these studies, 14 studies reported using non-zero priors, which 12 studies got the priors from previous studies that used the same questionnaire, one study mentioned the use of a pilot study to get priors and one study did not mention the method. Five studies used zero priors. Some studies reported using design generator software such as Ngene (n=6) or SAS® (n=7) to design their DCE (Supplementary file Table A 2.1.1).

Statistical analyses and software

The majority of the studies (n=23) used conditional logistic regression models. Two (106, 107) of these studies used rank-ordered logit in addition to conditional logit. Two studies used multinomial logit (124, 132). Five articles used random-effects probit models and two studies applied multinomial probit. Multi-level mixed effects logit, latent class method, generalized estimation equation (GEE) and conditional probit were each used in one study. One study mentioned a new model, the Zermelo-Bradley-Terry (ZBT) Model (133), to value health states.

Stata was the most commonly used software to analyse the data (n=18). SAS, R, Matlab and WinBUGs were other software used to analyse the DCE data (Supplementary file Table A 2.1.1).

Randomisation and consistency

Randomisation was usually done in a few ways; by randomly allocating respondents to a block of choice tasks (n=19), which was used in the majority of studies, or randomisation was achieved based on the order of choice sets or the question position among participants (n=13). In some studies, respondents were randomly assigned to choice-sets (n=5) (Supplementary file Table A 2.1.2).

Table 2.2 Consistency Methods

Number of studies	Method of consistency	The theory stated behind the method
9	**Dominant choices**	Using dominant choices among choice sets where one health state is better than all other states or is the worst state among all, or using duration as a dominant option, it means checking consistency with respondents who always chose the health profiles with the longest duration
6	**Best and worst health state**	Coefficients logically consistent if the worse health states has the least value and increases logically as the health state gets better (best health state has greatest value) (134)
7	**Question ordering**	Checking the answers to see whether the participant always selects the alternative that is on the left-hand side or on the right-hand side.

Methods for assessing consistency included consistency in respondents and consistency in coefficients, one study reported that it had consistency in responses (114) and five studies discussed consistency in coefficients but did not mention which method was used to check for this (115). In 12 studies, the authors did not report any consistency test.

Anchoring

All four methods for anchoring described by Rowen et al (104) were observed across the included studies as well as three additional methods not described by Rowen. The first method is to anchor using the dummy coefficient for dead, an approach used in only one study in this review (106). The second method is to anchor the worst state using TTO, again used in a single study (135). The third method is to map the DCE on to TTO, which was used in four papers (115, 136-138). The last approach is a hybrid of methods one and two used in two papers (130, 135). Two studies (104, 139) reported all anchoring methods in their study. The Bansback et al study (113) used the value of coefficients in the conditional logit for anchoring when an attribute for duration is included (DCEtto). The same method as Bansback's study, was used in 13 of the included studies. Two articles just mentioned anchoring between 0 and 1 for dead and full health respectively, and five studies anchored between -1 and 1, best and worst health state. One study (128) rescaled their utilities using an algorithm developed based on the SG method. Six studies did not report any anchoring method (Supplementary file Table A 2.1.2).

Quality assessment using the CREATE checklist

In general, the majority of the studies scored well against the CREATE checklist (117) with an average score of 86.5 percent (ranging from 68.4 to 100 percent). The item "sample size/power calculations are stated and rationalized" was the least observed item in the studies. The average score for subgroup health states values was the highest (average of 99 percent).

2.5 Discussion

A systematic review by de-Bekker Grob et al (119) showed that there were only two published studies that had used a DCE to value utility within the QALY framework between 2000 and 2010, and the review by Clark et al (110) showed that between 2009-2012, four studies had used DCEs to value utility within the QALY framework. In recent years, the use

of DCEs to estimate utility values has increased. This review addressed the methods when designing and analysing DCEs to value health states and outlined a variety of methods when developing a DCE for MAUIs. There is no apparent standardised method for design or analysis; however, there were similarities in the studies reviewed. For instance, the majority of studies used an online panel and a sample of the general population. There are divergent opinions on whose preference should be considered in valuation studies. On the one hand, some researchers believe that the general population should be the respondents as they are representative of the insured population (either via social or private insurance schemes) and therefore collectively they should dictate values. On the other hand, some researchers suggest that a person that has the disease can better value the health state and their opinion is more realistic (140). Maybe that it is not one or the other, rather, both are relevant but should be used for different purposes. This debate is ongoing for valuation studies of condition-specific MAUIs.

Variation in methodology and design might be considered useful by some researchers. It might promote learning about the impact of methodological choices and help guide decisions about future protocols or elements of new methodology that may be applicable to a wide range of DCE studies. However, while the variation in methods remains, it is important for decision-makers to understand the features and limitations of utility value sets in order to rely on these for decision making. Researchers need to continue to strive for value sets that are valid and reflect true preferences whilst continuing to educate and be guided by decision-makers and the public.

Design

Deciding which experimental design to use is an important step and has a key role when performing a DCE. Design defines the coefficients that can be estimated in analysis and can ensure the estimate is as precise as possible for a given sample size. As identified in our review

a range of methods have been used by studies in this area. Classical designs such as factorials use orthogonal designs, which apply orthogonal arrays employing different sources, in which the attributes are statistically independent and attribute levels occur equally (141). Using a full factorial design requires presentation of all combination of attributes; the Prosser et al (129) study used a full factorial design but excluded the dominated choices in the design. The authors state that decreasing the number of health states makes the design less efficient. Usually, there are numerous combinations, making presentation of all potential choices unfeasible; therefore, a fractional factorial design is generally more pragmatic to elicit respondent preferences.

In practice, designs which are near-orthogonal are usually also applicable, as the orthogonal designs usually exist for particular combinations. A near-orthogonal design, such as D-efficiency, is often used to measure the relative goodness of a fractional factorial design (142). Results of this review illustrate the trend towards using D-efficient designs in this area, which reflects a general trend observed in previous DCE methodology reviews (110, 119). The use of Bayesian efficient designs has also been increasing (136, 139) this might be due to the use of more random priors parameters instead of fixed priors parameters in the designs to obtain more stable designs. And in the Bayesian efficient designs, the priors are considered to be random parameters (143, 144).

In order to design the DCE the researchers can decide whether to include zero priors or non-zero prior information (expected values), there is controversy over using priors in the design. Additional work should be undertaken to understand the changes in the model when zero or non-zero prior information is included, which may then help to inform future studies.

Presenting the DCE choices set and training the participants about choosing health states is usually done by a brief introduction about the DCE at the beginning of the DCEs.

Adding a question to the checklist about the method of the presentation of the task within the DCE studies, might be beneficial, this can be an area for future work in DCE of MAUI.

Consistency

When designing the DCE, researchers need to consider strategies to identify and prevent high choice inconsistency. Regardless of how choices are presented to respondents, there should be logical consistency across questions. Logical consistency has a key role and is the main principle of choice theory in modern economics. Consistency describes logical ordering of health states in the method (145, 146), and it can be checked by the way the choices are presented to the respondents. For instance, changing the position of the alternatives or order of choice sets (111, 122) or checking if there are respondents who continuously chose the health profiles with the longest duration when duration is involved (147). However, it is difficult to claim that for a participant who always chooses the health state with the longest duration, that this might not actually be their 'true' preference. Estimated weights should increase when moving from worst health to best health state which can be easily assessed. 'True' preference means that there is no bias in choosing the health state and it is the participants actual preferred health state.

One of the main reasons for individuals responses to get deleted from DCE data sets is that the respondents are identified as having lexicographic preferences, meaning that the individual makes decisions by ranking attributes and consistently choosing the attributes in the order of their highest priorities; for example, the participant always chooses the option with the highest level of duration (148). However, based on the aim of DCE it may not be preferable for researchers to delete these respondents (149). For instance, some studies (127, 148) accomplished the analysis with the complete sample stating that deleting the individuals with dominant preferences will not change the estimated parameters, while the Bansback et al (113)

study deleted observations with lexicographic preferences, explaining that those individuals failed to understand the task. Although future studies could present the results with and without excluding participants and provide justification for selection of an approach, it remains unclear as to which set of results is more indicative of the true underlying preferences based solely on the data collected from the choice set. A valuable avenue for future research may be exploring the best method to check choice stability and consistency in choice studies and the extent to which apparently unstable or inconsistent choices can be identified and subsequently handled within the analysis.

Choice tasks

The de-Bekker Grob's (119) study discussed that the mean number of choices across all health-based DCES is 14, while Bridges et al (118) suggest that between 8 to 16 choices is a reasonable number in practice. Two studies stated that models using 10 choice pairs are more consistent than 15 pairs (137, 148). We found that the majority of the studies had between 8 to 12 choice tasks although the maximum number of choice sets was 32.

Having a large or a small number of choice sets has its own advantages and disadvantages. The advantage of having many choices is that it facilitates estimation of mixed logit models; however, if the number of choices is high, the participants might become fatigued and may not give complete attention to the questions and thereby decrease the completion rate (121). For instance, the Burr et al (114) study used 32 choices that caused the respondents to take a median of 75 minutes to answer the DCE questionnaire, and as a result, they could not include the individuals with severe case of the disease. Similarly, where the number of choices is high, participants may revert to heuristic choices, such as lexicographic preferences. In addition, including a low number, may result in a lack of information needed

for reliable analysis while higher tasks per person could provide one avenue to estimating individual preference algorithms.

Anchoring

Health state values can be anchored on a 0-1 scale where one presents the best health and zero represents dead (150), however using different descriptive systems, health states worse than dead are possible, which have a negative utility assigned to them (151). Studies (115, 128) that anchored health states using the best and the worst health states, where health states are scaled from –1 to +1, need to re-anchor the health states to estimate QALYs. Depending on the design of each study the anchoring might be different, yet the utilities need to be in the interval of zero (dead) to one (full health) scale if the results want to be incorporated into QALY calculations (152).

In DCEs, different models of anchoring health states to provide utility values have been applied. Usually, in DCE studies the health states are not compared to death directly and setting dead as zero is a less-direct approach compared to TTO and SG methods. The anchoring for DCE is done mostly by using a combination of DCE and a different but related preference based method such as TTO, or including duration as an option in the design, however, if a DCE is being applied to overcome the limitations of other preference based methods, using this method alone to anchor utility values is its own analytical challenges (153).

Based on the studies, we assume that anchoring the DCEs with duration (DCEtto), may be less complicated compared to other methods; however, there is still a debate regarding how to best present the duration attribute in DCE studies in the DCEtto approach (132). Further research is required to understand how duration may affect the selection of an item without

42

leading to any bias and more broadly, to determine the most appropriate way to anchor DCE data.

Analysis

The use of probit analyses has decreased from 1990 -2012 (110) while the use of logit analyses has risen in these years. The results of our study agree with this finding; recent papers mostly used logit analyses rather than probit analyses. There was no unique method and approach in all studies. However, most of the studies used conditional logit as their analysis method. The study by Hauber et al (154) provides a useful guide to explain the statistical methods to derive preference weights from DCEs, and choosing the best analysis method based on the aim and the data of their study.

2.6   Conclusions

Reflective of the results of this review, there is no gold standard DCE method. Given that there are different methods used to measure health-state values, future studies are required to compare various methods in order to indicate the best method based on the aspects of the value set. As the theoretical literature on DCEs evolves so too will the methods applied to analyse DCE data. It is evident however, that DCEs will continue as an approach to elicit preferences for valuing health states.

3   Chapter 3: Using Rasch and Factor Analysis to Develop a Proxy-Reported Health State Classification (Descriptive) System for Cerebral Palsy

STATEMENT OF CONTRIBUTION TO CO-AUTHOR PUBLISHED PAPER:

This chapter includes a published co-authored paper. The bibliographic details of the co-authored paper, including all authors, are:

Bahrampour M, Downes M, Boyd RN, Scuffham PA, Byrnes J. Using Rasch and factor analysis to develop a Proxy-Reported health state classification (descriptive) system for Cerebral Palsy. Disability and Rehabilitation. 2020:1-9. https://doi.org/10.1080/09638288.2019.1709565

Copyright statement: Copyright of this manuscript belongs to disability and rehabilitation Journal at Taylor & Francis publication. Authors permitted to use for scholarly purposes, personal website. Also authors can share their preprint anywhere at any time. Further details are available: https://authorservices.taylorandfrancis.com/research-impact/sharing-versions-of-journal-articles/

I contributed to the original idea of the study and undertook the data cleaning and data analysis. I also contributed to the interpretation of the results and writing up the manuscript and coordinated the response to reviewers.

(Signed) _____ (Date)_____

Mina Bahrampour

(Countersigned) _____ (Date)_____

Corresponding author of paper: Mina Bahrampour

(Countersigned) _____ (Date)_____

Supervisor: Prof. Paul Scuffham

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

Check for updates

Using Rasch and factor analysis to develop a Proxy-Reported health state classification (descriptive) system for Cerebral Palsy

Mina Bahrampour[a] , Martin Downes[a] , Roslyn N. Boyd[b] , Paul A. Scuffham[a] and Joshua Byrnes[a]

[a]Centre for Applied Health Economics, School of Medicine & Menzies Health Institute Queensland, Griffith University, Brisbane, Australia; [b]Queensland Cerebral Palsy and Rehabilitation Research Centre, Centre for Children's Health Research, The University of Queensland, Brisbane, Australia

ABSTRACT

Purpose: The Cerebral Palsy quality of life instrument is a well-known health-related quality of life measure for children with Cerebral Palsy. Due to its length it is not suitable as the basis of a preference-based instrument. The aim of this study is to develop a short version of the Cerebral Palsy quality of life instrument that can subsequently be scored as a multi-attribute utility instrument through assigning preference-based values.

Methods: A sample of 473 participants who have a child with Cerebral Palsy completed the Cerebral Palsy quality of life instrument(proxy-version) instrument. After deleting questions related only to the proxy, the dimensional structure was obtained using exploratory factor analysis. Extended Rasch analysis was then undertaken to test the psychometric performance of items and select the best item to represent each dimension. Expert opinion was sought to confirm the dimensions and items.

Results: A six-dimension classification system was identified, in which four domains were extracted from the factor analysis. Following expert opinion, two other domains were also added, as these were considered to have significant impact on health-related quality of life in children with Cerebral Palsy.

Conclusions: The combination of Factor and Rasch analysis along with consultation with patients, clinicians and experts in health-related quality-of-life instrument development, has resulted in a short version of the Cerebral Palsy quality of life instrument.

➤ **IMPLICATION FOR REHABILITATION**
- This study provides the first classification system for children with Cerebral Palsy.
- The Cerebral Palsy-six dimension (CP-6D) survey, which is a short version of Cerebral Palsy Quality Of Life instrument, can be timesaving when measuring quality of life in children with Cerebral Palsy.
- The short version (CP-6D) can be used in preference based measurement and generate quality adjusted life years for children with Cerebral Palsy.

Introduction

Cerebral palsy (CP) is the most common neurological disorder in children [1]. This life-long disability occurs due to a non-progressive defect or lesion in the maturing brain that happens during pregnancy or shortly after birth [2]. Globally, over 17 million people have CP with a prevalence of 2.0–3.5 per 1000 live births in developed countries [3]. In the US, CP is one of the top five most costly conditions on a per capita basis with an average Medicaid cost of $43,338 annually for a child with CP [4]. Cerebral Palsy is a permanent condition, with the majority of healthcare interventions aimed at reducing the long-term effects of this condition and improving a persons quality of life [5]. Due to limited resources, health care interventions and treatments need to be cost-effective to ensure maximum health outcomes are achieved [6,7].

Cost-utility analysis is one of the most common methods for evaluating new health care interventions. It is recommended by healthcare reimbursement agencies globally, including the National Institute for Health and Care Excellence (NICE) in the United Kingdom [8] and the Pharmaceutical Benefits Advisory Committee (PBAC) in Australia [9]. Cost Utility Analyses typically use quality-adjusted life years (QALY) as a summary health outcome measure [10] with the incremental cost per QALY ratio used to determine the value of new healthcare interventions [11].

Quality Adjusted Life Years combines both quality (morbidity) and quantity (mortality) of life within a single summary measure [12]. To produce QALYs, health utilities (weights) are required to capture the relative importance of living in different health states. Utility values can be obtained from existing multi-attribute utility instruments (MAUIs) [13], also known as preference-based measures (PBMs). Utility values are anchored on a scale where 1.0 represents full health and typically, 0.0 is death. PBMs have two elements: (1) a classification system, which is a set of dimensions and items that provide a description of all relevant states of health, and (2) a set of preference weights for each health state defined by the classification system. A classification system is required as health-related quality of life (HRQOL) measures are typically too long to undertake a meaningful valuation study.

Using Rasch and Factor Analysis to Develop a Proxy-Reported Health State Classification (Descriptive) System for Cerebral Palsy

3.1 Abstract

Purpose: The Cerebral Palsy quality of life instrument is a well-known health-related quality of life measure for children with Cerebral Palsy. Due to its length, it is not suitable as the basis of a preference based instrument. The aim of this study is to develop a short version of the Cerebral Palsy quality of life instrument that can subsequently be scored as a multi-attribute utility instrument through assigning preference based values.

Methods: A sample of 473 participants who have a child with Cerebral Palsy completed the Cerebral Palsy quality of life instrument (proxy- version) instrument. After deleting questions related only to the proxy, the dimensional structure was obtained using exploratory factor analysis. Extended Rasch analysis was then undertaken to test the psychometric performance of items and select the best item to represent each dimension. Expert opinion was sought to confirm the dimensions and items.

Results: A six-dimension classification system was identified, in which four domains were extracted from the factor analysis. Following expert opinion, two other domains were also added, as these were considered to have significant impact on health-related quality of life in children with Cerebral Palsy.

Conclusions: The combination of Factor and Rasch analysis along with consultation with patients, clinicians and experts in health-related quality-of-life instrument development, has resulted in a short version of the Cerebral Palsy quality of life instrument.

Keywords: Cerebral Palsy, Health related quality of life, Health state classification system, Rasch analysis, Item response theory

46

3.2 Introduction

Cerebral palsy (CP) is the most common neurological disorder in children (3). This life-long disability occurs due to a non-progressive defect or lesion in the maturing brain that happens during pregnancy or shortly after birth (155). Globally, over 17 million people have CP with a prevalence of 2.0 - 3.5 per 1,000 live births in developed countries (156). In the US, CP is one of the top five most costly conditions on a per capita basis with an average Medicaid cost of $43,338 annually for a child with CP (26). Cerebral Palsy is a permanent condition, with the majority of health care interventions aimed at reducing the long-term effects of this condition and improving a person's quality of life (157). Due to limited resources, health care interventions and treatments need to be cost-effective to ensure maximum health outcomes are achieved (158, 159).

Cost-utility analysis is one of the most common methods for evaluating new health care interventions. It is recommended by health care reimbursement agencies globally, including the National Institute for Health and Care Excellence (NICE) in the United Kingdom (38) and the Australian Pharmaceutical Benefits Advisory Committee (PBAC) in Australia (39) .Cost-Utility Analyses typically use quality-adjusted life years (QALY) as a summary health outcome measure (31) with the incremental cost per quality-adjusted life years ratio used to determine the value of new health care interventions (49).

Quality Adjusted Life Years combines both quality (morbidity) and quantity (mortality) of life within a single summary measure (160). To produce QALYs, health utilities (quality-adjusted life year weights) are required to capture the relative importance of living in different health states. Utility values can be obtained from existing multi-attribute utility instruments (MAUIs) (97), also known as preference based measures (PBMs). Utility values are anchored on a scale where 1.0 represents full health and typically, 0.0 is death. PBMs have two elements:

1) a classification system, which is a set of dimensions and items that provide a description of all relevant states of health, and 2) a set of preference weights for each health state defined by the classification system. A classification system is required as HRQoL measures are typically too long to undertake a meaningful valuation study.

The cerebral palsy quality of life instrument (CPQOL) is a condition-specific measure designed to measure the HRQoL in children and adolescents with CP. This instrument has been used widely in clinical trials for children with CP in different countries (161-166) and shown to be reliable and valid in this population (167). The CPQOL has 65 questions (or items), which is burdensome to collect data. The existing scoring algorithm of this instrument uses a 0-100 scale for quality of life with all questionnaire items equally weighted; for instance, improvement in reducing the amount of pain is weighted the same as an improvement in the child's ability to play alone.

As attributes do not have an equal effect on HRQoL, the current scoring system is inadequate in reflecting what matters most. Moreover, as the scoring system does not reflect the potential magnitude that people are willing to forego in order to be in one health state relative to another, utility values, that are necessary for Cost-utility analysis, cannot be determined (168).

Utility values previously applied within cost-effectiveness studies of interventions for children and adolescences with CP have been sourced from generic MAUIs, which are mostly designed for adults, such as the European quality of life five dimensional (EQ-5D) (169), Health utility index (HUI) (58, 170-174) and Assessment of Quality of Life (AQoL) (174). Arguably, these MAUIs may not be suitable nor sensitive for use in children with CP (58). Generic MAUIs have limited ability to capture all relevant and fundamental dimensions of HRQoL that may characterise a given condition, for instance for CP (168, 175). In addition, generic MAUIs

might lack sensitivity to differences or changes in health states for specific conditions. Thus, condition-specific MAUIs have been developed. Examples include: Asthma (176), Cancer (177), and Multiple Sclerosis (178). Given the specific but diverse manifestation of CP among patients, a condition-specific preference based MAUI is required.

The aim of this study is to develop a cerebral palsy specific health state classification system (sometimes referred to as a descriptive system) for the CPQOL-child (proxy version). The classification system is the first step in developing a preference based measure that can be used in economic evaluations for interventions designed to improve the health outcomes of children with CP.

3.3 Materials and methods

CPQOL Instrument

The CPQOL has been developed to measure the quality of life in children with CP. This instrument has three versions: child report, proxy report and a teen version (62, 63). The CPQOL instruments assess the following domains: patients' social wellbeing and acceptance (12 items); physical health (11 items); feelings about functioning (12 items); emotional well-being and self-esteem (6 items); and pain and impact of disability on wellbeing (8 items). The proxy report version evaluates two extra domains which are relevant to the parent/proxy quality of life, including access to services (12 items) and family health (4 items). Item responses are reported on a 9 point scale, with descriptors provided for five levels; very unhappy, unhappy, neither happy nor unhappy, happy and very happy, in which one is the worst and nine is the best (179). The CPQOL-child parent/proxy report was used in this study, which originally contains 65 items. This version has been designed to evaluate HRQoL of children with CP by gathering information from their parents or guardian. The parent/proxy version, has shown good reliability and internal consistency (64). As the aim of this study is to construct a CP

specific health state classification based on items common in all versions of the CPQOL and specific to the child/adolescents HRQoL, the items belonging to the two domains on access to services and family member's health, have been excluded from the analysis. Consequently, the 49 items common across all versions are included in this analysis.

CPQOL Datasets

Three data sets of CPQOL responders were used in this study. These data were gathered from studies that administered the CPQOL-child parent/proxy to estimate the child's HRQoL (180-182). The studies included are: "INCITE: A randomised trial comparing constraint induced movement therapy and bimanual training in children with congenital hemiplegia" (n= 236) (181), the "Move it to improve it (Mitii)" study (n= 104) (180), and the "Australian Cerebral Palsy Child Study" (n= 133) (182). In these studies, the CPQOL was completed by parents to capture their child's perceptions of their quality of life. The children's ages ranged between one and 17 years of age.

Classification System Development

A three-step process was used to develop the classification system, namely: 1) domain selection, 2) item selection; and 3) item response level reduction. This is consistent with the approach outlined by Brazier et al (183).

Domain selection

Factor Analysis

Data are required to be tested for suitability prior to constructing a factor analysis, therefore data were screened using the Kaiser-Meyer-Olkin (KMO) index (184), which shows the proportion of variance in the variables, has a threshold of greater than 0.5, and Bartlett's

test of Sphericity (185) that has a threshold of p<0.05. A p-value less than 0.05 on Bartlett's test indicates that individual variables are sufficiently correlated for a factor analysis to be performed (186, 187).

Exploratory factor analyses (EFA) with oblique factor extraction (Promax) was conducted to find the number of dimensions (factors). An oblique rotation was used which assumes that the factors are related (188). The number of factors was determined and preserved using multi-extraction criteria based on eigenvalues, parallel analysis, and scree plots. An eigenvalue cut-off value of one was used to establish the number of factors (189) and the scree plot was visually checked to find the bend-point in the data and identify where the curve flattens out. Where results differ when using eigenvalues and scree plots independently, multiple factor analyses were constructed to find the optimized factor structure.

The items in each factor were chosen based on the factor's uniqueness and item loadings. Greater uniqueness (1- communality) shows the weaker relevancy of the items to the factors, so the selected criteria for preserving the item's uniqueness was ≤ 0.5. Hence, if the item uniqueness was more than 0.5 then the item was deleted from the factors. In addition, items were removed from the model if they had a loading of less than 0.3 on any factor (190). The analyses were repeated until all items satisfied these pre-specified criteria.

The key dimensions of the original instrument are important, yet factor analysis may produce models that do not contain the structure of the original instrument completely; in addition, when selecting items to be included in the classification system, it is important that these items are clear and amenable to the valuation study and cover all the important domains that might affect the quality of life. Moreover, as psychometric evidence is not the only criteria for item selection (191), an expert panel was also established to assess domain relevance. This team, to choose the best possible domain structure, included three health economists, an

epidemiologist, four clinicians and a patient representative. One of the clinicians on the team was involved in the original development of CPQOL.

Item selection

Rasch analysis

To select the most representative item for each dimension, item response measurement by extended Rasch analysis was conducted for each identified factor. Rasch analysis is a mathematical technic and belongs to a class of Item Response Theory (IRT) models. Compared to classical test theory, Item Response Theory is a modern approach to test the reliability of psychological test assessment, assessing how and why a person responds to an item. Both methodologies can be used to assess individual change in a clinical context. However, when estimating measurement precision Item Response Theory and classical test theory may differ. Classical test theory assumes that it will be equal for all participants regardless of their attribute level, but Item Response Theory assumes that it depends on a latent-attribute (unobservable) value. This difference leads to results with potentially different conclusions regarding statistical significance (192). Rasch analysis produces a set of items, which are selected from a survey, to conduct a scale in which the item difficulties and person measures are determined (193).

In this study, the Rasch Partial Credit Model (PCM) was used, as this model permits multilevel item responses for all items. Final item reduction and selection was accomplished based on item loading in the factor analysis and item fit criteria in the extended Rasch analysis. Items were eliminated if they did not meet key criteria (described below). The same expert panel that was used for domain selection was also used to assess item relevance and clarity leading to the selection of a single item suitable for each dimension. Final item selection was

based on more than 80 percent agreement among the experts, which generally infers reasonable reliability (194).

Item level reduction

If disordering is observed in the item levels after extended Rasch analysis, item levels were collapsed and item level reduction will be undertaken. We were constrained to the wording of the original questionnaire the CPQOL levels because this questionnaire is on 5 level scale.

Test item performance

Factor analysis was undertaken using Stata (version 14) statistical software, and extended Rasch analysis and goodness of fit, were done using R software (version 3.4.3 eRm package). Item fit statistics were also used to investigate the overall fit to the extended Rasch model. A significant Chi-square (χ^2) fit statistic indicates misfit of the model where the significant level for this analysis was considered to be 0.05. Where misfit was identified, items were removed because of multidimensionality until the extended Rasch model statistics showed an acceptable fit which was defined with respect to Item-fit and Person-fit values close to zero. Other indices that were used to interpret the goodness of fit for items were in-fit (inlier-pattern-sensitive fit statistic) and out-fit (outlier-sensitive fit statistic) measures, which identify valid fitting if the square of their means is within the range of 0.5 to 1.5.

Differential item functioning (DIF) was also considered as it can also affect model fit and generalisability of the classification system to all patients. Differential item functioning occurs when different groups in the sample (between genders for example) systematically answer an item differently. This can be assessed by measuring the item response for subgroups. In this study, the DIF was estimated with respect to gender, age (school-aged or under), and

condition severity (Gross Motor Functioning Classification System, GMFCS; and Manual Ability Classification System, MACS). Gross Motor Functioning Classification System and Manual Ability Classification System both have five levels in which one is the lowest severity and five is the high severity (195). Person separation reliability was also estimated, which is a reliability index similar to Cronbach's alpha (α).

3.4 Results

Among all 473 pooled responses to the CPQOL, 56% were male, 51.2% had an age less than or equal to 5 years, 89.5% had a GMFCS and level of less than or equal to II (mild CP), and 90.39% had a MACS level of less than or equal to II (mild CP) (Table 3.1).

Table 3.1 demographic statistics (%)

Variables	Percent (%) N=473
Gender Female Male	44.0 56.0
Age ≤ 5 (non- school) > 5 (school)	51.2 48.8
GMFCS level ≤ 2 (mild) > 2 (moderate and severe)	89.5 10.5
MACS level ≤ 2 (mild) > 2 (moderate and severe)	90.4 9.6

**Missing: age (n=6), Gross Motor Functioning Classification System (GMFC) (n=14), Manual Ability Classification System (MACS) (n=15)

Factor and Rasch analysis

Results of the Kaiser-Meyer-Olkin index and Bartlett's test of Sphericity test show that the data are suitable for factor Analysis (p-value < 0.001 and Kaiser-Meyer-Olkin index = 0.94). The Exploratory factor analyses results showed that five factors had eigenvalues higher than one and the item structure based on the factor analysis was different from the original instrument. The scree plot was bent on the third factor showing that three factors could be

extracted, and parallel analyses predicted four factors. To find the best factor structure to fit the data, the factor analysis was run with three, four, five and six factors, where the item "*sleep*" was excluded. This item (sleep) did not load on any factor but showed a high uniqueness, suggesting a different construct being measured by this item compared to other items. After input from the expert panel, this item was used as an individual domain. The optimized factor analysis (defined as no loadings under 0.3 and reduced cross-loadings) was identified as the four-factor extraction. Another item, "*use of hand*", was added as it was considered an important aspect of CP based on consultation with the patient representative. Based on interpretation from the expert panel, a total of six domains were included in the classification system, the six domains included were considered: "Social wellbeing and acceptance", "Physical health", "Communication", "Pain and discomfort", "Manual ability" and "Sleep".

The summary statistics of item loadings and dimensions are provided as a supplementary file. The results show that the overall amount of missing data varied across items, while most of the items did not have any missing data or was less than 1%, the item "*the amount of pain*" had the highest missing rate (5.35%). All items were free from floor effects yet some of the items exhibited ceiling effects (more than 20% of responses) including *sleep* (25.90%) (Appendix A 3.1).

The items for each domain identified from the Factor analysis were used in extended Rasch analysis with a Rasch model fitted for each of the four domains. From 14 items that represented the domain social wellbeing and acceptance, nine items were removed after assessing the Rasch model goodness of fit, so the remaining number of items representing factor one reduced to five items. Three items were removed in the physical health domain after Rasch analysis and nine items remained. The other two domains, communication and pain and

discomfort did not change, leaving five items in the communication domain and three items in the pain and discomfort domain.

Table 3.2 illustrates the goodness of fit for each dimension based on extended Rasch analysis results, in which all show a good fit to the extended Rasch analysis. Mean item fit, mean person fit and person separation index are used to show the goodness of fit for each dimension. All indices show the model has a good fit and dimensions are suitable for use.

Table3.2 Rasch model goodness of fit for each dimension

Dimension	Number of items	Mean Item fit	Mean Person fit	Person separation reliability
Social wellbeing and acceptance	5	-0.001	0.006	0.951
Physical health	9	-0.010	-0.011	0.932
Communication	5	0.004	0.043	0.906
Pain and discomfort	3	0.007	0.004	0.900

An item by item psychometric and extended Rasch analysis for the four multi-item dimensions are presented in Table 3.3. Items were considered to have fitted the model well as indicated by the chi-square, p-values and in-fit mean squares. For all factors, items that demonstrated different item functioning as identified by age, gender and disease severity (using GMFCS ans MACS) were not considered for the final health state classification system.

The extended Rasch analysis identified the item accepted by people in general had a higher p-value and covered a broader range compared to the item getting along with adults. Thus, accepted by people in general was the item chosen to represent the social wellbeing and acceptance domain.

Based on the p-value, range and DIF results, the item the way they get around was chosen for the domain physical health. This item was considered to appropriately and broadly represent this domain by the expert panel.

For the communication domain, two items fitted well, their ability to participate in social events outside of school and the way people communicate with them. The latter item however, was considered better in measuring the main characteristics of this domain based on input from the patient representative on the expert panel who attested that "how people communicate with children with CP is an important factor and it impacts their wellbeing" and that this item may also be more generalizable for children not attending school, hence it was selected in the final classification system.

For the domain Pain and discomfort, two items fitted well, how much pain do you have and feelings about the amount of pain. Like most items in the original instrument were based on feelings, and how the quality of life is measured using well-being, the item feelings about the amount of pain was included in the classification system based on the input of the expert panel.

Table 3.3 Results of the Rasch analysis for items in domains

	p-value (χ^2)	In-fit MSQ	Item range	DIF
Factor one: social wellbeing and acceptance				
get along with other children outside of school	0.613	0.988	-2.967, 0.791	GMFSC, Age
get along with adults	0.974	0.872	-2.991, 0.593	-
accepted by other children outside of school	0.986	0.827	-2.725, 0.923	gender
accepted by adults	1.000	0.619	-3.096, 0.396	GMFCS, MFCS, Age
accepted by people in general	1.000	0.647	-3.028, 0.693	-
Factor two: Physical health				
ability to participate at school	0.817	0.883	-3.318, 0.963	-
ability to participate in recreational activities	0.988	0.865	-2.880, 0.943	Age
participate in sporting activities	0.374	0.951	-2.501, 1.104	Age
their physical health	0.191	1.038	-3.447, 0.755	Age
the way they get around	1.000	0.694	-4.147, 0.968	-
the way they look	0.853	0.994	-3.481, 0.468	Age
their ability to keep up physically with their peers	0.999	0.779	-3.273, 1.662	-
feels about themselves	0.935	0.914	-2.844, 0.567	-
the way they use their legs	0.211	0.998	-3.166, 1.279	-
Factor three: communication				
their ability to participate in social events outside of school	1.000	0.765	-2.437, 0.775	-
their ability to participate in your community	1.000	0.687	-2.773, 0.966	GMFSC
the way they communicate with people they know well	0.807	1.018	-2.895, 0.039	-
the way they communicate with people they don't know well	0.386	0.963	-2.600,1.005	Age
the way people communicate with them	1.000	0.590	-2.516, 0.744	-
Factor four: Pain and discomfort				
How much pain does your child have	1.000	0.569	-0.163, 2.441	-
the amount of pain they have	1.000	0.604	-0.098, 2.006	-
How much discomfort does your child experience	0.878	0.851	-0.681, 2.636	-

The overall item reduction process is shown in Figure 3.1.
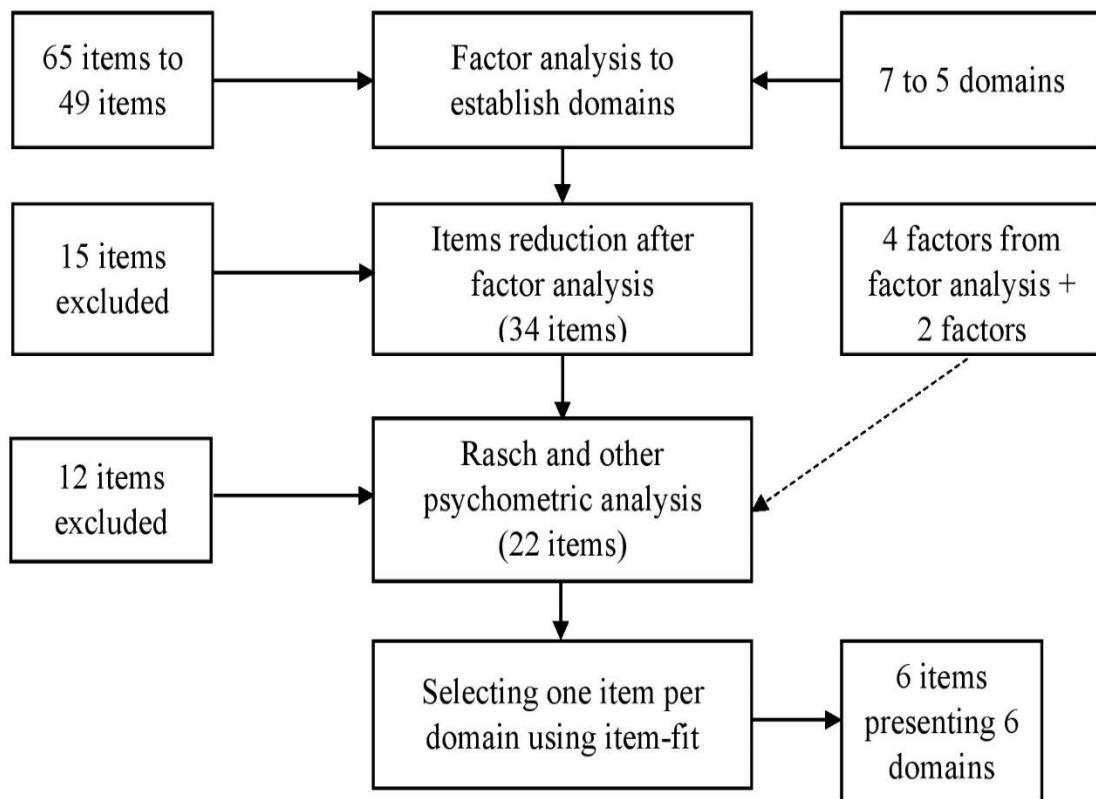
Figure 3.1 CPQOL item reduction process

As it is shown in Figure 3.1, the beginning of this process there were seven domains with 65 items in the questionnaire, two domains and their related items were excluded that resulted in five domains with 49 items. After factor analysis, 34 items within four domains proceeded to Rasch analysis. Subsequently, 22 out of 49 were remained after the extended Rasch analysis and fitted the model. Final item selection for each domain was completed based on consensus with the expert panel. The resulting classification system, including the two extra domains as described above, was comprised of six domains and six items representing those domains.

Item response reduction was undertaken as modelling with nine levels showed disordering in the extended Rasch analysis. Hence, through consensus among the expert panel the nine level response scale was reduced and a five-level classification system was adopted. This conclusion was reached in part due to the fact that level descriptions for most of the CPQOL items were in five levels, so no new response-level definitions were needed except for the amount of pain item which only was classified in "not at all upset" and "very upset", the three other levels were decided by the experts.

The final classification system comprises six items that each represents a conceptual domain of CP HRQoL: social wellbeing and acceptance, physical health, communication, pain and discomfort, sleep and manual ability. Each item has five levels scaled from very unhappy to very happy to reflect the original wording of the CPQOL instrument. (Table 3.4). In order to test the validity of the newly developed 6 item questionnaire, a reliability test was done using a sub sample and found that the new instrument, hereafter termed the CP-6D, is correlated with the original instrument with a correlation of 0.89.

Regarding the terminology used in the paper and the classification, the items within the CPQOL instrument ask parents "how do you think your child feels about…?", so for instance when we talk about the item "the way they use their hands", this term has been adopted from the question asking parents "how do you think your child feels about the way they use their hands?"

Table 3.4 The proposed classification system for the new preference based measure (CP-6D)

Dimension	Description
Social well-being and acceptance	They feel **very unhappy** about how they are accepted by people in general They feel **unhappy** about how they are accepted by people in general They feel **neither happy nor unhappy** about how they are accepted by people in general They feel **happy** about how they are accepted by people in general They feel **very happy** about how they are accepted by people in general
Physical health (motor functioning)	They feel **very unhappy** about the way they get around They feel **unhappy** about the way they get around They feel **neither happy nor unhappy** about the way they get around They feel **happy** about the way they get around They feel **very happy** about the way they get around
Communication	They feel **very unhappy** about the way people communicate with them They feel **unhappy** about the way people communicate with them They feel **neither happy nor unhappy** about the way people communicate with them They feel **happy** about the way people communicate with them They feel **very happy** about the way people communicate with them
Pain and discomfort	They are **very upset** about the amount of pain they have They are **moderately upset** about the amount of pain they have They are **somewhat upset** about the amount of pain they have They are **slightly** about the amount of pain they have They are **not at all upset** about the amount of pain they have
Sleep	They feel **very unhappy** about how they sleep They feel **unhappy** about how they sleep They feel **neither happy nor unhappy** about how they sleep They feel **happy** about how they sleep They feel **very happy** about how they sleep
Manual ability	They feel **very unhappy** about the way they use their hands They feel **unhappy** about the way they use their hands They feel **neither happy nor unhappy** about the way they use their hands They feel **happy** about the way they use their hands They feel **very happy** about the way they use their hands

3.5 Discussion

This study describes the process of developing a health state classification system for children with CP patients from the CPQOL-child parent/proxy version. It is the first study jointly applying a factor analysis and extended Rasch analysis to the CPQOL to define the construction of a CP-specific health state classification system. A valid classification system is the first step of developing a CP-specific preference based measure by identifying an instrument that can be valued using preference elicitation methods. Moreover, this is the first study to develop a condition-specific classification system for CP. The next step will be to perform a valuation study to develop a utility set using general population preferences that can be used in economic evaluation of interventions in CP.

The rationale for selecting the CPQOL-child (parent/proxy version) as the basis for the development of a CP-specific preference based measure is that the CPQOL is a valid and frequently used outcome measures to evaluate the quality of life in children with CP. This is primarily based on the psychometric and clinical data among quality of life instruments that have been used in populations of children with CP (61, 167). The classification system designed in this study, which we are coining as the Cerebral Palsy 6 Dimension (CP-6D), is a six-item health state classification system derived from the CPQOL instrument, covers the most important and relevant HRQoL domains for people with CP including sleep, pain, communication, physical health, social well-being and acceptance and manual ability (15, 196-198). Moreover, validity testing demonstrated that the CP-6D is highly correlated with the original instrument.

In this study, sleep was added as a domain, even though none of the previous quality of life outcome measures used in this population had included this aspect. Sleeping disorders are more frequent in children with CP than naturally developed children (199, 200). However, it must be noted that our analysis was based on parent / proxy reporting and sleep of a child can disturb the sleep of a parent.

Additionally, the original CPQOL-child (parent/proxy version) instrument did not have '*communication*' as a domain as items regarding communication were previously a priori subsumed in a domain of feelings about functioning. However, after the factor analysis, it was demonstrated that communication is a domain on its own right. Moreover, this is consistent with other CP quality of life outcome measures (61) such as the PedsQL CP module (201), caregiver priorities and index of life with disabilities (CPCHILD) (202), DISABKIDS (203), and care and comfort hypertonicity questionnaire (C&CHQ) (204).

Another domain was also included, manual ability based on the (item using hands). Although it did not load as an individual domain in the factor analysis, this ability is considered an important characteristic of the condition and indicates the extent to which a child needs support (205). The manual ability item, however, did highly load on the domain physical health, yet the item of best fit for the physical health domain was getting around. This may be due to the fact that the item getting around covers a broader spectrum compared to using hands when estimating physical health in the CP HRQoL. Cerebral Palsy severity is clinically classified into three categories, which includes gross motor functioning, manual ability and communication. As such, the research team decided to include manual ability as its own domain to reflect established clinical classification.

Developing a preference based measure from an existing instrument has its own advantages and disadvantages. A limitation of deriving a classification system from an existing HRQoL measure, as compared to developing a "de novo" instrument, is that researchers are confined to using the items and wording of the original instrument (206). The advantage of deriving a classification system from an existing instrument is that the utility values derived from the new condition-specific instrument can be used with existing data sets, and in future studies using the CPQOL (183). At the same time, the classification system could be used instead of the long questionnaire in clinical trials and other studies where previously the CPQOL would have been used. The CP-6D can be used in clinical trials instead of the CPQOL, which will be less time consuming, and the results can be used to calculate QALYs for economic evaluations. However, the framing of the questions based on who will answer the questions (proxy-report or self-report), are different and would require change dependent on the respondents.

Like all research, this study has limitations. First, whilst CP is a multifaceted disability with many comorbidities, the classification developed here is framed with respect to well-being ("feelings about …"). This is an artefact of the development and wording of the CPQOL and is not inclusive of factors specific to function limitations. The CPQOL instrument is however, the most widely used instrument in clinical trials for children with CP for measuring the quality of life and has shown to be valid in this population (64, 207, 208). In addition, because all the questions are framed with respect to wellbeing, we did not use "emotional wellbeing" as a domain because, even though it was a domain in the original instrument it was not identified from the factor analysis. If emotional wellbeing was to be included as a domain, it would more than likely be interpreted as an overall question given that the response levels of all other items are framed with respect to happiness. Despite this, subsequent validation of the classification system may be warranted. In addition, most of the respondents in our study were parents of children with mild CP, which arguably limits the applicability of this analysis to more severe patients. However, severe patients were included in our data and items that exhibited DIF analysis, including based on the severity of the disease, were removed. Moreover, all analyses herein are with respect to the proxy-report version of the CPQOL to develop a classification system. However, it is unknown to what extent the same concluding classification system would have been developed where the data based on the self-report version. This is an interesting area of research more broadly with respect to measuring and valuing patient health-related quality of life. Future studies may well be warranted to identify how well the CP-6D developed here works for the self-report version of the CPQOL. This paper has used Australian sample to develop the classification system, for future research it might be useful to do the analysis on a different setting using the same method.

In conclusion, this study provides the first classification system for children with CP. It has been developed based on the most widely used CP specific HRQoL measure, the

CPQOL-child (parent/proxy version). The new classification system, Cerebral Palsy-six dimension (CP-6D), contains six dimensions each with one item and each item comprised of five response levels. This is an important landmark in the development of a CP specific multi-attribute utility instrument for use in economic evaluations. This tool can be used in Future valuation studies to calculate QALYs in the economic evaluation of treatments and interventions for people with CP.

Ethical approval: This article does not contain any studies with human participants performed by any of the authors.

4 Chapter 4: Developing a cerebral palsy-specific preference based measure for a six-dimensional classification system (CP-6D): protocol for a valuation study

<u>**STATEMENT OF CONTRIBUTION TO CO-AUTHOR PUBLISHED PAPER:**</u>

This chapter includes a published co-authored paper. The bibliographic details of the co-authored paper, including all authors, are:

Bahrampour M, Norman R, Byrnes J, Downes M, Scuffham PA. Developing a cerebral palsy-specific preference-based measure for a six-dimensional classification system (CP-6D): protocol for a valuation study. BMJ open. 2019; 9(9):e029325.

Copyright statement:

I contributed to the original idea of the study and planning the data collection activities data analysis and writing up the manuscript. I coordinated the response to reviewers.

(Signed) _____ (Date)_____

Mina Bahrampour

(Countersigned) _____ (Date)_____

Corresponding author of paper: Mina Bahrampour

(Countersigned) _____ (Date)_____

Supervisor: Prof. Paul Scuffham

BMJ Open

Developing a cerebral palsy-specific preference-based measure for a six-dimensional classification system (CP-6D): protocol for a valuation study

Mina Bahrampour,[1] Richard Norman,[2] Joshua Byrnes,[1] Martin Downes,[1] Paul A Scuffham[1]

Check for updates

[1]Centre of Applied Health Economics, Griffith University, Menzies Health Institute Queensland, Brisbane, Queensland, Australia
[2]School of Public Health, Curtin University, Perth, Western Australia, Australia

Correspondence to
Mina Bahrampour;
mina.bahrampour@griffithuni.edu.au

ABSTRACT

Introduction Cerebral palsy (CP) is a lifelong condition. The CP quality of life (CPQOL) instrument is a frequently used disease-specific instrument to assess health-related quality of life (HRQoL) in people with CP, but it cannot be used to generate quality-adjusted life years (QALY) which are the basis of cost utility analysis (CUA). Generic utility instruments (such as the EQ-5D or SF-6D) that are used to value HRQOL may be insensitive to small but important health changes in children with CP. This study aims to generate a preference-based scoring algorithm for the CP six dimensions (CP-6D), a classification system developed from the CPQOL.

Methods and analysis A discrete choice experiment with duration (DCEtto) will be administrated to value health states described by the CP-6D classification system. These health states will be presented to members of Australian general population and parents of children with CP via an online survey. Conditional logit regression will be used to produce the utility algorithm for CP-6D.

Ethics and dissemination The Griffith University Human Research Ethics Committee approved for the study (reference HREC/number 2018/913). The developed algorithm can be applied to previous and future economic evaluation of interventions and treatments targeting people with CP which have used either the CPQOL or CP-6D.

INTRODUCTION

Health economic evaluations are widely used to value interventions, treatments, procedures and policies. The most prevalent methods in economic evaluation are cost-effectiveness analysis (CEA) and cost utility analysis (CUA),[1] where CUA can be considered a special case of CEA.[2] CEA compares the costs and outcomes resulting from an intervention relative to one or more comparators and estimates the incremental cost of each additional unit of outcome. CUA builds on this by using a summary health outcome measure that considers both mortality and quality of life, usually through the quality-adjusted life years (QALY).[3] CUA is intended to enable comparison across disease areas and

Strengths and limitations of this study

► Developed a rigorous protocol to undertake the valuation of the CP-6D.
► Utility values generated from this study, using the DCE method, will be estimated for all interventions and treatment for people with CP in which CPQOL or CP6D have been applied.
► The DCEtto method will help anchoring the preference value sets on a 0–1 scale for both general population and people with CP.
► The difference in preferences between the general population and people familiar with CP will be compared.

hence to be of more use to policymakers. For this reason, CUA is widely recommended as the primary method of economic evaluation by healthcare reimbursement agencies.[4]

The QALY allows comparison of different health interventions or treatments over a time interval. To estimate QALYs resulting from an intervention, health utility weights (or health state values) are required. These utility values are attributed to health states, typically described by a classification system, and are anchored on a scale from 0 to 1, where 0 represents dead and 1 represents full health.[5] Health state values are usually generated from preference-based instruments that use preference elicitation techniques such as time trade-off (TTO), standard gamble (SG) and discrete choice experiment (DCE)[6] to assign a value to each health state described by the underlying classification system.

Preference-based instruments can be either generic-specific or condition-specific.[7] Conventionally, generic instruments have been more widely used, as they allow estimation of QALYs based on consistent health state classifications and utility data across all conditions and treatments. However, as these

67

Developing a cerebral palsy-specific preference based measure for a six-dimensional classification system (CP-6D): protocol for a valuation study

4.1 Abstract

Introduction: Cerebral palsy (CP) is a life-long condition. The Cerebral Palsy Quality Of Life (CPQOL) instrument is a frequently used disease-specific instrument to assess Health-Related Quality of Life (HRQoL) in people with CP, but it cannot be used to generate quality adjusted life years (QALY), which are the basis of cost-utility analysis (CUA). Generic utility instruments (such as the EQ-5D or SF-6D) that are used to value HRQoL may be insensitive to small but important health changes in children with CP. This study aims to generate a preference based scoring algorithm for the Cerebral Palsy 6-dimension (CP-6D), a classification-system developed from the CPQOL.

Methods and analysis: A Discrete Choice Experiments with duration (DCEtto) will be administrated to value health states described by the CP-6D classification system. These health states will be presented to members of Australian general population and parents of children with CP via an online survey. Conditional logit regression will be used to produce the utility algorithm for CP-6D.

Ethics and dissemination: The Griffith University Human Research Ethics Committee approved for the study (reference HREC/number 2018/913). The developed algorithm can be applied to previous and future economic evaluation of interventions and treatments targeting people with CP that have used either the CPQOL or CP-6D.

Keywords: Choice experiment, Cerebral palsy, health related utility values, Preference

Strengths and limitations of this study

- Developed a rigorous protocol to undertake the valuation of the Cerebral Palsy 6 Dimension (CP-6D).

- Utility values generated from this study, using the DCE method, will be estimable for all interventions and treatment for people with CP in which CPQOL or CP6D have been applied.

- The DCEtto method will help anchoring the preference value sets on a 0-1 scale for both general population and people with CP.

- The difference in preferences between the general population and people familiar with CP will be compared.

4.2 Introduction

Health economic evaluations are widely used to value interventions, treatments, procedures, and policies. The most prevalent methods in economic evaluation are cost-effectiveness analysis (CEA) and cost-utility analysis (CUA) (97), where CUA can be considered a special case of CEA (209). CEA compares the costs and outcomes resulting from an intervention relative to one or more comparators, and estimates the incremental cost of each additional unit of outcome. CUA builds upon this by using a summary health outcome measure which considers both mortality and quality of life, usually through the Quality Adjusted Life Years (QALY) (210). CUA is intended to enable comparison across disease areas and hence to be of more use to policymakers. For this reason, CUA is widely recommended as the primary method of economic evaluation by health care reimbursement agencies (211).

The QALY allows comparison of different health interventions or treatments over a time interval. To estimate QALYs resulting from an intervention, health utility weights, (or health

state values) are required. These utility values are attributed to health states, typically described by a classification system, and are anchored on a scale from zero to one, where zero represents dead and one represents full health (212). Health state values are usually generated from preference based instruments that use preference elicitation techniques such as time trade off (TTO), standard gamble (SG) and discrete choice experiment (DCE) (213) to assign a value to each health state described by the underlying classification system.

Preference based instruments can be either generic or condition-specific (183). Conventionally, generic instruments have been more widely used, as they allow estimation of QALYs based on consistent health state classifications and utility data across all conditions and treatments. However, as these instruments are designed to be generic, they may lack the ability to capture changes in important aspects of some health conditions, particularly for those smaller generic instruments which may consist of only a handful of domains by which to describe health. Therefore, there is a growing trend towards condition-specific instruments, which are likely to be more reflective of all important domains of health and more sensitive to changes in health-related quality of life in some diseases or disabilities (214).

There are two key elements of a preference based instrument: 1) a classification system, which provides a system for the full spectrum of health defined by a discrete number of health states; and 2) a utility value set, which provides a value for each health state as described by the classification system (183). Recently, a new health state classification system, the Cerebral Palsy 6-dimension (CP-6D) derived from the CPQOL (Cerebral Palsy Quality of Life) instrument has been developed (215). The CPQOL is designed to measure health-related quality of life in children and adolescents with CP (216). For developing the CP-6D, the parent/proxy version was used as there was a wide range of age groups from one to 17 year olds in the trials that administrated CPQOL; however, the survey was adjusted based on the

common questions in all versions. To develop the CP-6D statistical methods, including factor and Rasch analyses, were used. These methods were applied to identify the dimensions and the item to represent each dimension for the CP-6D. The dimensions of the CP-6D are social wellbeing and acceptance; physical health; communication; pain and discomfort; manual ability; and sleep.

The development of the second element of this preference based instrument, the health state value set, is the focus of the present study. However, there is considerable debate regarding who should determine the value health of each health state (90), as it is believed that people with the condition and the general population may value the dimensions of health differently (217). Conversely, and perhaps more traditionally, it has been argued it is the general population preferences that should be considered and used to determine the value of the utility health states. This line of argument is supported by two considerations. Firstly, it is the general population who are the taxpayers (or insurance fund members) and it is their money that will be used to fund new interventions and treatments, so it is ultimately this broader population base whose preferences should be considered. Secondly, the general population would be representative of all diseases and conditions and therefore the valuation might be considered more objective (218).

The goal of this study is to generate a preference based scoring algorithm for the CP-6D. To measure utility values of the health states defined by the CP-6D classification system, an algorithm will be developed using statistical models. This paper describes the methodology that will be used to collect data on preferences and develop the utility values for the CP-6D. When complete, the algorithm will enable data collected from any study using of the CPQOL to be used in the economic evaluation of treatments and interventions for people with CP by

converting responses from the CPQOL to a HRQoL utility value necessary for deriving QALYs.

Aims

1) To value health states generated from the classification system (CP-6D) with both a sample of the general population and a sample from people with CP registered with Queensland CP, using a discrete choice experiment (DCE);

2) To determine the difference between general population and people with CP health state utility values.

4.3 Methods and analysis

Preference and elicitation methods

Both ordinal and cardinal preference based elicitation methods can be used for preference elicitation. Standard Gamble (SG) and Time trade off (TTO) are the most widely used cardinal methods to value health states. SG and TTO are both choice-based approaches. The SG task asks participants to choose between a certain health state (the health state being valued) and a scenario with a probability (p) of gaining the best possible health outcome (full health) and a probability (1-p) of the worst possible health outcome (usually immediate death) (75). For health states better than dead, TTO asks individuals about how many years of living in full health followed by an immediate death they would be willing to trade compared to living in a health state for a definite number of years. It determines that the individual is willing to accept living in a full health state for fewer years relative to being sick for more years. However, it has been argued that risk aversion and time preference can bias an individual's response using the SG and TTO respectively (106, 107). Additionally, it has been shown that these tasks are highly complex (219) and sensitive to mode of administration (76), hence there has recently been a shift towards ordinal methods, such as the discrete choice experiment (DCE) approach

(220). DCE has gained popularity in health economic research since it uses ordinal response to estimate interval measures (99) and can potentially allow for more flexible characteristics of the utility function (147).

In this study, DCE with duration (DCEtto) will be used to elicit preference for health states generated and described by a new CP specific instrument (CP-6D). The DCE model asks participants to choose their preferred health state over hypothetical alternatives (221). However, the values derived from a standard DCE (which compares health states without duration) are not anchored on a utility scale from 0 - 1, and therefore cannot be used to measure QALY directly. The values generated from the DCEtto however, can directly anchor their relative preference onto the utility scale with the inclusion of a duration attribute (113, 222).

The CP-6D domains (physical health, social well-being and acceptance, communication, pain, sleep and manual ability) each consist of five levels. In this study, the choice sets for the DCE will be generated from levels of these six domains with the addition of one attribute representing duration. The ordinal levels for all dimensions (except Pain and duration) are: very unhappy, unhappy, neither unhappy nor happy, happy, and very happy. The ordinal levels for the pain domain are: very upset, moderately upset, somewhat upset, slightly upset and not at all upset (Table 4.1). These levels are derived from the wording of the original CPQOL instrument.

Table 4.1 The CP-6D classification system (215)

Dimension	Description
Social well-being and acceptance	They feel **very unhappy** about how they are accepted by people in general They feel **unhappy** about how they are accepted by people in general They feel **neither happy nor unhappy** about how they are accepted by people in general They feel **happy** about how they are accepted by people in general

	They feel **very happy** about how they are accepted by people in general
Physical health	They feel **very unhappy** about the way they get around They feel **unhappy** about the way they get around They feel **neither happy nor unhappy** about the way they get around They feel **happy** about the way they get around They feel **very happy** about the way they get around
Communication	They feel **very unhappy** about the way people communicate with them They feel **unhappy** about the way people communicate with them They feel **neither happy nor unhappy** about the way people communicate with them They feel **happy** about the way people communicate with them They feel **very happy** about the way people communicate with them
Pain and discomfort	They are **very upset** about the amount of pain they have They are **moderately upset** about the amount of pain they have They are **somewhat upset** about the amount of pain they have They are **slightly** about the amount of pain they have They are **not at all upset** about the amount of pain they have
Sleep	They feel **very unhappy** about how they sleep They feel **unhappy** about how they sleep They feel **neither happy nor unhappy** about how they sleep They feel **happy** about how they sleep They feel **very happy** about how they sleep
Manual ability	They are **very unhappy** about the way they use their hands They are **unhappy** about the way they use their hands They are **neither happy nor unhappy** about the way they use their hands They are **happy** about the way they use their hands They are **very happy** about the way they use their hands

The duration attribute contains five levels (1, 3, 5, 7, 10 years). The upper limit of 10 years was used as it is the commonly used as the fixed period of life years lived in a less than full-health health state in TTO valuations (223).

DCE tasks

A full factorial design is a combination of all attributes and their levels. Health states included in the DCE tasks will be generated using a combination of levels across dimensions, which will involve a mixture of high levels for some dimensions and low levels for other dimensions. The combination of attributes and levels for a full fractional in this study (six dimensions of the CP-

6D each with 5 levels) would however result in $5^6 = 15,625$ health states and when adding duration would result in 78,125 (5^7) health state profiles and over a billion (78,125*78,125) possible pairwise combinations of any two health state and duration combinations. Each pair-choice presents two scenarios in which the respondents are asked to choose their preferred health state to live until they die. An example of the DCE pairwise task is presented in Table 4.2.

Table 4.2 An example of a Discrete Choice experiment choice pair

Which health state do you prefer?

Domain		
Social wellbeing and acceptance	You feel **happy** about how you are accepted by people in general	You feel **very happy** about how you are accepted by people in general
Physical health	You feel **happy** about the way you get around	You feel **unhappy** about the way you get around
Communication	You feel **unhappy** about the way people communicate with you	You feel **very happy** about the way people communicate with you
Pain and discomfort	You are **slightly upset** about the amount of pain you have	You are **very upset** about the amount of pain you have
Sleep	You feels **very happy** about how you sleep	You feel **very unhappy** about how you sleep
Manual ability	You are **very unhappy** about the way you use your hands	You are **happy** about the way you use your hands
Duration	Living in this health state for **3 years** and then die	Living in this health state for **1 year** and then die
Which health state do you prefer?	**Health State A** ☐	**Health state B** ☐

Given the number of health profiles, it would not be appropriate to present all combinations to participants. As such, a practical subset of health states will be selected (reduced number of health states) and used in the experiment whilst optimising the efficiency of the design. Specifically, a D-efficient design will be used to increase the efficiency of data collection.

In experimental design, choices should be selected that can examine both main effects (the effect of each independent variable on the dependent variable) and possible interactions (preference for an attribute based on the level of another). However, in a DCEtto the disutility of levels of the instrument through interactions with duration would also be measured. For this study, a design will be developed using the design generator software Ngene (224), the design will be based on D-efficiency criteria to select pair-wise choice sets. The design will be generated to capture the two factor interactions involving duration with duration anchoring the DCEtto on a scale of full health (1.0) and dead (0.0). The DCE will be designed without any priors for the pilot study, after that the priors generated from the pilot study will be applied to the final design.

It has become standard to ask each respondent to complete between 8 -12 choice sets when using DCE to value health states generated from a multi-attribute instrument. This range has been considered to maximise data collection per respondent without incurring significant responder bias such as to undermine the quality of the data. In this study, 12 choice sets per responder was chosen. The full range of choice sets will use a block design with 12 choice sets in 20 blocks so as to obtain responses with respect to 240 health states from both the general population and for people with CP. To prevent order bias by respondents the sequence of choice pairs will be randomized within each block. The blocking will help the balance in the levels of attributes (225), and will also ensure that the number of respondents per block is equal.

Survey

Participants

In this study, a sample from the general population in addition to the parents/proxies of a child with CP sample will be recruited. Therefore, a value set for each group can be presented and also the preference differences between the two groups can be obtained.

General population and CP registry group

An online survey will be administrated to an Australian general population from May 2019. The survey contains several sections, the beginning section is an introduction to the research and the respondents will be asked to provide consent in order to continue with the rest of the survey. After the person accepts to be a part of the study, the next section will require participants to provide demographic data (age, gender, education, income, and health status using the AQoL-4D (226), which will allow a determination of whether the sample is representative of the Australian population. Next, the participants will be asked to describe their own health state using the CP-6D, with the next section containing the DCEtto tasks (i.e. 12 plus an addition choice sets).

 In the beginning of each section, there will be an introduction and a guide on how to complete the questions. Consistency will be checked using one of the choice tasks asked twice in the DCEtto and at the end of the DCEtto tasks the individuals will be asked to rate the difficulty of the questions on a scale from one to four.

Specifically, for the CP population demographic questions will also include the level of Gross Motor Function Classification System (GMFCS) (227) instrument and the Manual Ability Classification System (MACS) [29] instrument. These two scales describe the severity of the disease and CP functioning. The GMFCS, which is based on an individual's movement, is a multi-level categorization technique that has five levels specifying the rate of how much help

a person with CP needs and identifies whether the person needs a wheelchair or can walk independently (227). The MACS describes how much assistance the person needs to use their hands (228, 229). Both tools are widely used for the purpose to capture the child's functional level, in CP studies (230)

Sample size and recruitment

The respondents will be recruited from an existing Australian online panel administrated by Survey Engine (231), which is a survey company with expertise in online DCEs. Each respondent will be paid a small amount as to complete the survey (approximately AUD$10 each). The respondents will be anonymous, and only de-identified data will be provided to the researchers. The respondents will be presented a web-link to access the survey and this will able them to complete the tasks at their convenience. An online panel is a cost effective way to recruit respondents, and has been widely used in general population valuation studies (232).

Statistical efficiency is a key focus of experimental design, and is a major factor in determining necessary sample size. In similar work, a sample size of 1,000 to 2,000 has been demonstrated to produce small confidence intervals, even if the experimental design is not maximally efficient (233). Based on this, a sample of 2,000 people from the general population will be recruited. A study by Lancsar and Louviere stated that more than 20 respondents per choice set is required to estimate reliable models (234). In our study using a sample of 2,000 individuals means there will be more than 20 respondents per choice set which is consistent with the Lancsar and Louviere study.

With respect to the sampling of the Australian population with experience of CP, parent/proxy of children with CP will be recruited from an existing Australian CP registry (ACPR). Based on previous research conducted in this population an estimated number of people with CP who can be recruited from this registry is n=300. For this population, a proxy (e.g. parents or

guardian)(235) will be asked to completed the survey, as the child may lack the cognitive ability to complete all tasks within the survey. This is consistent with other studies in which a proxy has been employed to complete the tasks (236).

Pilot study

The survey will be soft-launched using a sample of 100 from the general population and 30 people from the CP population. The piloting will start in May 2019 starting with the CP group. If no changes are made, the full data collection will occur including data collected prior and subsequent to the soft launch. The first dataset will be collected to:

 1) Pilot the classification system using parents/proxies opinions for validation of the 6 domains;

 2) Check question difficulty, clarity and understanding by individuals. To this end, two questions will be added asking the participants how difficult they found the question on a scale of one to four. (Difficulty in answering and difficulty in understanding the question);

3) Indicate the feasibility of the duration levels;

4) Assess the time spent by each individual on the whole survey and for each DCE task. This will be used to determine participant burden and the extent to which response behaviour changes during the survey.

5) Determine the functioning of the whole survey. The pilot study will show practical issues when completing the tasks and indicates if any revisions are required including if the block design and randomisation of task ordering are operating as designed.

Patient and Public Involvement

No patient involved

Analytical plan

To estimate health state values and determine the coefficients for main effects and interactions between the main effects, regression models will be fitted. This will include the conditional logit, as outlined by McFadden (82), and mixed logit to potentially account for correlation of error terms in individual respondents (237, 238) (That is the likely correlation among the multiple responses (i.e. choices) provided from each individual). The final model selection will be determined based on model fit where model fit will be assessed with respect to log likelihood ratio chi-square or Mcfadden's pseudo R square (237).

DCEtto is based on random utility theory (RUT), which states that the utility value of an attribute in a scenario can be explained by both fixed and random components (239). As such, using the coefficients from the best-fitted model, values for the health states of the utility-based instruments can be estimated. The scoring algorithm will be developed using the model introduced by Bansback et al (113) where an extra attribute for duration is included in our design. Algorithms will be developed to convert responses for the quality of life instrument to utility-based instruments based on the coefficients of the selected models. The data will be analysed using Stata and R.

Potential incomplete data will be explored and further decision will be made based on the type of missing data. Probably a multiple imputation will be done, however, the data need to be gathered first.

The health state values resulted from this study can be used in to calculate QALYs in economic evaluation of treatments and interventions for people with CP, where CPQOL has been used.

4.4 Ethics and dissemination

There are no known health or safety risk associated with participants in any aspects of the study. The DCE opening however, there is a consent form which allows the participants to voluntary enter the survey and to ensure that they will be fully informed about the aim of the study.

This study will estimate utility values for the new CP specific instrument (CP-6D). This would be the first CP specific preference based instrument to value utility for people with CP from the Australian population. The algorithm developed from this study can be used to generate health state values for any study that has previously used CPQOL or plans to do so in the future, as the CP-6D was derived from CPQOL. The utility values generated from the algorithms developed in the present study can then be used to estimate QALYs for cost-utility analyse of new treatment or interventions that are aimed for people with CP. The distribution of the results of this study will be through publication in academic journals and presenting in conferences.

5 Chapter 5: Utility Values for the CP-6D a Cerebral Palsy Specific Multi Attribute Utility Instrument using Discrete Choice Experiment

STATEMENT OF CONTRIBUTION TO CO-AUTHORED PUBLISHED PAPER

This chapter includes a published co-authored paper. The bibliographic details of the co-authored paper, including all authors, are:

Bahrampour M, Norman R, Byrnes J, Downes M, Scuffham PA. Utility Values for the CP-6D, a Cerebral Palsy-Specific Multi-Attribute Utility Instrument, Using a Discrete Choice Experiment. Patient (2020). https://doi.org/10.1007/s40271-020-00468-x

I contributed to the original idea of the study and undertook the data collection, cleaning and analysis. I also contributed to the interpretation of the results and wrote up the manuscript and coordinated the response to reviewers.

(Signed) _____ (Date)_____

Mina Bahrampour

(Countersigned) _____ (Date)_____

Corresponding author of paper: Mina Bahrampour

(Countersigned) _____ (Date)_____

Supervisor: Prof. Paul Scuffham

ORIGINAL RESEARCH ARTICLE

Utility Values for the CP-6D, a Cerebral Palsy-Specific Multi-Attribute Utility Instrument, Using a Discrete Choice Experiment

Mina Bahrampour[1] · Richard Norman[2] · Joshua Byrnes[1] · Martin Downes[1] · Paul A. Scuffham[1]

Abstract

Background and Objective The CP-6D is a new preference-based measure derived from the CPQOL, a cerebral palsy-specific quality-of-life questionnaire. The CP-6D contains six dimensions, each with five levels. A preference-based value set is required to score the CP-6D on a utility scale and render it suitable for cost-utility analysis. This study aims to estimate the utility value set for the CP-6D for interventions for people with cerebral palsy (CP).

Methods A discrete choice experiment was designed and administrated to an adult Australian online panel. Each respondent answered 12 choice sets. Each choice was presented as a combination of the health state from the CP-6D and duration spent in that health state before death. Conditional logit and mixed logit regression were used to analyse the data. The utility values were estimated as a ratio of the coefficient of each dimension to the coefficient of the duration.

Results A total of 2002 participants completed the survey and responded to each choice. Generally, the dimension levels were monotonic, meaning the coefficients reflected the ordered nature of the levels in each dimension. The dimensions relating to manual ability, social well-being and acceptance had the greatest effect on choice. The value of the worst 'pits' health state is -0.582.

Conclusion This study provides the first CP-specific utility value set that can potentially be used in cost-utility analyses of interventions for people with CP where the CPQOL has been applied, both prospectively and retrospectively.

Key Points for Decision Makers

There is no utility value set for cerebral palsy (CP); this study provides the first value set for the CP-6D, a new preference-based multi-attribute utility instrument derived from the widely used cerebral palsy quality of life questionnaire, CPQOL.

The reimbursement process in many countries is informed by cost-utility analysis (CUA). The availability of a value set for CP-6D assists the application of CUA for CP interventions.

✉ Mina Bahrampour
mina.bahrampour@griffithuni.edu.au

[1] Centre for Applied Health Economics, School of Medicine, Menzies Health Institute Queensland, Griffith University, Nathan, QLD, Australia

[2] School of Public Health, Curtin University, Bentley, WA, Australia

1 Introduction

Multi-attribute utility instruments (MAUIs) are health-related quality-of-life (HRQoL) measures that allow the valuation of health on a scale of one (representing full health) to zero (representing dead) [1]. These instruments may be generic and applicable to any condition, or be condition-specific. The relative merits of the two approaches are well known. A generic instrument maximises the comparability of results across conditions. However, it may be insensitive to certain differences in health states that manifest in ways not well captured by a generic instrument's dimensions. Conversely, a condition-specific instrument is likely to be sensitive and minimise item redundancy in that specific population, but concern can then be raised about the generalisability of findings, which is important for the conduct and interpretation of cost-utility analysis (CUA) [2]. Research has been undertaken to estimate the health state utility values from condition-specific instruments; however, there is still concern as to whether condition-specific measures can be used alone in economic evaluation [3–5].

MAUIs have two components, the classification system, which can be developed de novo or derived from an existing quality-of-life instrument, and the utility algorithm (also

△ Adis

83

Utility Values for the CP-6D a Cerebral Palsy Specific Multi Attribute Utility Instrument using Discrete Choice Experiment

5.1 Abstract

Background and objective: The CP-6D is a new instrument derived from the CPQOL, a cerebral palsy-specific quality of life questionnaire. The CP-6D contains six dimensions, each with five levels. A preference based value set is required to score the CP-6D on a utility-scale and render it suitable for cost-utility analysis. This study aims to estimate the utility value set for the CP-6D for interventions for people with cerebral palsy (CP).

Methods: A discrete choice experiment was designed and administrated to an adult Australian online panel. Each respondent answered 12 choice sets. Each choice was presented as a combination of the health state from the CP-6D and duration spent in that health state before death. Conditional logit and mixed logit regression were used to analyse the data. The utility values were estimated as a ratio of the coefficient of each dimension to the coefficient of the duration.

Results: A total of 2002 participants completed the survey and responded to each choice. Generally, the dimension levels were monotonic meaning the coefficients reflected the ordered nature of the levels in each dimension. The dimensions relating to manual ability, social well-being and acceptance had the greatest effect on choice.

Conclusion: This study provides the first CP-specific utility value set, which can potentially be used in cost-utility analysis of interventions for people with CP where the CPQOL has been applied.

Keywords: CP-6D, Discrete Choice Experiment, Quality of Life, Valuation, Cost-Utility Analysis, Economic Evaluation, Cerebral Palsy

Key points

- There is no utility value set for cerebral palsy; this study provides the first value set for the CP-6D, a new preference based multi-attribute utility instrument derived from the widely used cerebral palsy quality of life questionnaire, CPQOL.

- The reimbursement process in many countries is done by cost-utility analysis (CUA). The availability of a value set for CP-6D assists the application of CUA for CP interventions.

Declaration

5.2 Introduction

Multi-attribute utility instruments (MAUIs) are health-related quality of life (HRQoL) measures that allow the valuation of health on a scale of one (representing full health) to zero (representing dead) (240). These instruments may be generic and applicable to any condition, or be condition-specific. The relative merits of the two approaches are well known. A generic instrument maximises the comparability of results across conditions. However, it may be insensitive to certain differences in health states that manifest in ways not well captured by a generic instrument's dimensions. Conversely, a condition-specific instrument is likely to be sensitive and minimise item redundancy in that specific population, but concern can then be raised about the generalisability of findings, which is important for the conduct and interpretation of cost-utility analysis (CUA) (60). Research has been undertaken to estimate the health state utility values from condition-specific instruments; however, there is still concern as to whether condition-specific measures can be used alone in economic evaluation (241-243).

MAUIs have two components, the classification system, which can be developed de novo or derived from an existing quality of life instrument, and the utility algorithm (also known as the value set) (183). Typically, preference based methods are used to generate utility values for MAUIs. Preference methods vary and include cardinal methods, such as the standard gamble (SG) and time trade-off (TTO), and ordinal techniques, such as discrete choice experiments (DCE), best-worst scaling (BWS) and other ranking exercises. The use of DCEs to value health states has increased; in this study, we use a DCE to value health states within the CP-6D (69, 79).

The DCE is a preference based elicitation technique that asks individuals to choose between health states over a set of different hypothetical scenarios to elicit their preferred health state (244). It can measure the relative weights of various attributes embodied within

health states (245). If the duration of the health state is also included in the DCE tasks, the method is often called DCEtto, and this approach allows anchoring of health states onto a full health-dead utility-scale. The resulting utility values can be used to estimate quality adjusted life years (QALYs) (246, 247).

Cerebral palsy (CP) is the most common physical disability during childhood (2) and covers a set of clinical features resulting from a non-progressive defect or lesion in the developing brain (1, 3, 4). Although associated with development, it is a life-long disability with high costs (20). The focus of this work is the CPQOL, a pre-existing CP-specific quality of life (QoL) instrument that is widely used to measure QoL in the CP population (167). The Cerebral Palsy 6-dimension (CP-6D) is a health state classification (descriptive) system developed from the CPQOL (215). Its development is described elsewhere, but briefly, it was developed using Factor and Rasch analysis from a sample of 473 people with CP. It was shown that reducing the number of items of larger instrument does not introduce a significant loss of information; the CP-6D has a correlation of 0.89 with CPQOL. As described below, it contains six dimensions, each with five levels. (56).

A CP-specific utility value set is not available but would allow better quantification of health change to inform the economic evaluation of interventions and programs for people with CP. Thus, this study aims to generate a preference based scoring algorithm for the CP-6D that can be used in the economic evaluation of interventions for people with CP.

5.3 Methods

Health states in the DCE were described using the CP descriptive system (CP-6D). To develop the CP-6D, the CPQOL parent/proxy version was modified for use by everyone with CP. To allow conversion from the CPQOL, the levels of the CP-6D range from 'very happy' to 'very unhappy' in all the dimensions except pain, which was described from 'not at all upset'

to 'very upset'. Levels were described using the exact wording of the original instruments to allow direct translation from self-reported health to QALY scores (Table 5.1).

Table 5.1 Cerebral palsy 6-dimension (CP-6D) health state classification system (215)

Dimension	Description	Levels
Social well-being and acceptance	How you feel about the way you are accepted by people in general	very happy happy neither happy nor unhappy unhappy very unhappy
Physical health	How you feel about the way you get around	very happy happy neither happy nor unhappy unhappy very unhappy
Communication	How you feel about the way about the way people communicate with you	very happy happy neither happy nor unhappy unhappy very unhappy
Pain and discomfort	How you feel about the amount of pain you have	not at all upset slightly upset somewhat upset moderately upset very upset
Sleep	How you feel about how you sleep	very happy happy neither happy nor unhappy unhappy very unhappy
Manual ability	How you feel about the way you use your hands	very happy happy neither happy nor unhappy unhappy very unhappy

Valuation task

Different anchoring methods can be used to value health states using DCEs (69). In this study, the valuation task was based on the DCEtto approach of Bansback et al. (246) due to

its wide use in the field (69, 79). The DCE tasks involved two hypothetical health states selected from the CP-6D, each combined with a duration. The choice of pairs (over triples or quads) was based on minimising the cognitive burden on respondents, and the fact that pairs are more widely used in the relevant literature (79). The participants were required to choose their preferred health states between 'health state A' and 'health state B'. The valuation task comprised 12 different pairs of choice sets. A 13th pair was added as a repeat of the sixth pair to check consistency. Therefore, 12 different DCE tasks were asked in the survey. The duration levels were 1, 3, 5, 7 and 10 years; these were selected to align with other DCEtto surveys and the typical application of the TTO (248). Consequently, health states were presented with a total of seven dimensions (Table 5.2).

Table 5.2 A sample of a discrete choice experiment valuation task

Please choose your preferred health state		
How you feel about how you get accepted by people in general	Very Happy	Happy
How you feel about the way you get around	Happy	Very Unhappy
How you feel about using your hands	Unhappy	Happy
How you feel about the way people communicate with you	Neither happy nor Unhappy	Unhappy
How you feel about the amount of pain you have	Very Upset	Not at all Upset
How you feel about how you sleep	Very Happy	Happy
You will stay in this state for the following length of time, and then die	10 years	7 years
Please tick your preferred health state.	○ Health State A	○ Health State B

The DCE design

A total of 5^6 (i.e., 15,625) possible health states can be obtained from the CP-6D. When adding duration, there are $5^7 = 78,125$ possibilities and over a billion possible pairwise combinations of any two health-states and durations. Given the number of health states, it was inappropriate to present all combinations. Therefore, a practical subset of health states was selected and used in the experiment to optimise the efficiency of the design. Ngene software (version 1.2.1) was employed to design choice tasks, applying a D-efficient design to increase the efficiency of data collection. The design was dummy coded with interactions with duration.

The experimental design was applied to select 240 choice sets. To generate the final design, a pilot study with 80 participants was undertaken to estimate prior values for each level of each attribute in the survey. The pilot study was conducted with zero priors. Analysis of the pilot data revealed some inconsistent ordering and non-monotonicity in the levels of the health domains. This could be because the pilot was conducted with a small number of respondents. The non-monotonicities were removed by combining the disordered levels, as done in previous studies (112, 245, 246, 249). After the priors from the pilot study were finalised, the design method from Ngene was replicated but with non-zero priors to generate the final design.

To limit respondent bias, the sequence of choice pairs was randomised within blocks, and the number of respondents per block was balanced. Blocks were also applied to help the balance in the levels of attributes. The full range of choice sets was 13 choice sets in 20 blocks to obtain responses with respect to the Australian general population.

Data collection

Online survey administration was conducted by SurveyEngine, a company with expertise in online DCEs. Potential participants, who were members of an online panel of

respondents willing to answer research surveys, received an email that included an invitation to complete the survey and a hyperlink that allowed them to access it. To ensure the sample was representative of the Australian population, recruitment was conducted using a quota control cut-off approach based on age and gender: once the quota for each age and gender category was reached, respondents were no longer accepted. The respondents could complete the survey at their leisure and, upon completion, received a small amount paid by the panel administrators (approximately $10).

Survey structure

The first page of the survey displayed a brief explanation of CP and informed the participants that the survey is used to develop health state values to help understand CP-6D. The second page of the survey contained consent information outlining the task in broad terms and describing the respondents' rights. The participants agreed to take part in the study by following the link to the main survey. The survey started with a brief guide on what a DCE is and how to complete the DCE questions. It consisted of the 13 choice pairs followed by a series of socio-demographic questions. Time spent on each page of the survey was automatically collected. At the end of the questionnaire, there were two difficulty questions concerning how difficult the respondent found it to understand the task, and how difficult it was to answer the questions on a scale from one to four (with one being easy and four being difficult).

We explored two of the more common patterns of lexicographic preferences in our data. First, we identified respondents who picked only option A or only option B for all choice pairs. Second, we identified respondents who always selected the longer duration. As further validation, we identified the proportion of respondents who failed the consistency check. There was not *a priori* choice to exclude those with seemingly lexicographic or inconsistent choices; rather, it is potentially useful to exclude them for data robustness.

Utility valuation

Data were analysed using the statistical software Stata 14 *clogit* and *mixlogit* commands. A two-factor interaction term was applied between the duration and attribute levels to estimate the utility values for the DCE data. Conditional and mixed logit were both used to check and distinguish the better-fit model by AIC and BIC criteria (250).

For the conditional logit, we used the *vce(cluster)* option in Stata to adjust standard errors to reflect repeated measures. The use of conditional logit for this type of analysis was outlined by McFadden (82). The econometric specification for the conditional logit in the context of a DCEtto reflects the assumptions of the QALY framework, specifically that the value of a health profile is a product of the value of health state and the duration spent in that state. We used the notation from Norman et al (121), where the utility of individual i choosing option k in choice set j is assumed to be:

$$Eq(1) \qquad U_{ijk} = \alpha DUR_{ijk} + \beta \acute{X}_{ijk} DUR_{ijk} + \varepsilon_{ijk}$$

Where α is the utility associated with the duration attribute (life years); β is the utility value associated with the level of each dimension in $\acute{X}_{ijk}$, which is the dummy variable presenting the levels of CP-6D. $\acute{X}_{ijk}$ contains 6 dimensions $\times$ (5-1) levels, resulting in 24 terms. This model presented every decrement from full health in each dimension within $\acute{X}_{ijk}$. The error term (ε) has a normal distribution (0, σ). To generate the utility value for each level, its coefficient (β) was divided by the coefficient of duration (α). Stata's *wtp* command (delta method) was used to estimate the confidence intervals (251).

The mixed logit was applied to examine choice preference heterogeneity in the sample population. This method was also adopted by King et al (112). Compared to Eq (1), this equation includes the individual deviations around the mean preferences:

$$Eq(2) \qquad U_{ijk} = (\alpha + \gamma_i)DUR_{ijk} + (\beta + \eta_i)\acute{X}_{ijk}DUR_{ijk} + \varepsilon_{ijk}$$

The α and βs in Eq (2) present the mean preference in the population, while the new variables γ_i and η_i are the individual variation from mean preference. Mixed logit also measures the standard deviation for the vectors γ_i and η_i.

Estimating health states

After estimating the coefficients and the decrement from full health for each dimension, to calculate, for instance, the utility value for health state 152241, the formula is 1 – (utility decrements of acceptance level 1, physical health level 5, manual ability level 2, communication level 2, pain level 4 and sleep level 1). This means that a person is very happy about how people accept them, very unhappy about getting around, happy about using their hands, happy about communicating with people and moderately upset about the amount of pain they have and very happy about how they sleep.

Ethics approval for this study was given by the Griffith University Human Research Ethics Committee, reference number: 2018/930.

5.4 Results

Sample characteristics

Overall, 4,752 people logged in to participate; 1,127 were immediately excluded for being over quota in age or sex, and 1,623 people did not complete the survey. A total of 2,002 people were eligible and completed the online survey. The demographic attributes of these people are presented in Table 5.3. The results indicate that the sample is broadly representative of the Australian population in age and gender, although they have higher stated education level.

Table 5.3 Socioeconomic characteristics of the sample and comparison with Australian population

		Study population(N)	Study population (%)	Australian population (%)
Gender	Male	976	48.75	49.03
	Female	1026	51.25	50.97
Age (years)	18-29	419	0.21	0.21
	30-39	362	0.18	0.19
	40-49	347	0.17	0.17
	50-59	327	0.16	0.16
	60-69	273	0.14	0.13
	70 and above	274	0.14	0.14
Education	Year 11 or below	250	0.12	0.28
	Year 12	354	0.18	0.17
	Trade certificate	287	0.14	0.24
	Diploma	319	0.16	0.09
	Bachelor's degree	539	0.27	0.14
	Postgraduate degree	253	0.13	0.09
Health insurance	Yes	1169	58.39	51
	No	833	41.61	49

Australian sex and age distribution (Australian Bureau of Statistics) from https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3101.

Utility estimation

The results of the conditional and mixed logit are presented in Table 5.4. In both methods, all coefficients were negative as expected, except level 2 of Physical health, which was not statistically significantly different from zero (p-value = 0.39). Moving from the best level to the worst, the coefficient values increased (in absolute terms), reflecting that the respondents' preferences were monotonic and consistent with the structure of the CP-6D. We chose conditional logit for the final value set because an economic evaluation holds greater interest in the mean response and the difference between society choices (98); in this study, preference heterogeneity was a secondary concern (112).

Table 5.4 Conditional logit and mixed logit results

	Conditional logit	Mixed logit
	Coefficients (standard error)	**Coefficients (standard error)**
Duration	0.446(0.018) ***	0.689(0.028) ***
Social wellbeing and Acceptance × Duration		
2	-0.001(0.005)	-0.008(0.007)
3	-0.059(0.005) ***	-0.089(0.008) ***
4	-0.095(0.006) ***	-0.152(0.009) ***
5	-0.119(0.007) ***	-0.188(0.01) ***
Physical health × Duration		
2	0.004(0.004)	0.001(0.007)
3	-0.035(0.005) ***	-0.048(0.007) ***
4	-0.085(0.005) ***	-0.134(0.008) ***
5	-0.111(0.006) ***	-0.167(0.009) ***
Manual ability × Duration		
2	-0.016(0.005) ***	-0.02(0.007) ***
3	-0.05(0.005) ***	-0.074(0.008) ***
4	-0.101(0.006) ***	-0.152(0.009) ***
5	-0.131(0.007) ***	-0.197(0.01) ***
Communication × Duration		
2	-0.009(0.004) **	-0.014(0.007) **
3	-0.028(0.004) ***	-0.041(0.007) ***
4	-0.081(0.006) ***	-0.123(0.008) ***
5	-0.097(0.006) ***	-0.144(0.008) ***
Pain and Discomfort × Duration		
2	-0.018(0.005) ***	-0.039(0.007) ***
3	-0.029(0.005) ***	-0.051(0.007) ***
4	-0.043(0.005) ***	-0.076(0.007) ***
5	-0.134(0.006) ***	-0.222(0.01) ***
Sleep × Duration		
2	-0.012(0.005) ***	-0.009(0.007) ***
3	-0.041(0.005) ***	-0.06(0.007) ***
4	-0.085(0.006) ***	-0.12(0.008) ***
5	-0.113(0.006) ***	-0.171(0.01) ***
Log-likelihood	-17,357	-16,711
AIC	34,765	33,523
BIC	34,986	33,966

The coefficients were estimated as the interaction of that level with duration.

The utility decrements for the levels of each dimension are presented in Table 5.5 with a 95% confidence interval. The larger utility decrements were allied with manual ability, and social well-being and acceptance. At the worst case (level 5) pain had the highest utility decrement.

Table 5.5 Utility decrements used in the CP-6D utility algorithm

Dimensions	Level	Utility decrement	Confidence interval (95%)	
Social wellbeing and Acceptance				
	1	0.000		
	2	-0.002	-0.022	0.018
	3	-0.132	-0.152	-0.112
	4	-0.214	-0.234	-0.193
	5	-0.267	-0.288	-0.247
Physical health				
	1	0.000		
	2	0.000	-0.011	0.028
	3	-0.077	-0.097	-0.058
	4	-0.191	-0.211	-0.172
	5	-0.249	-0.268	-0.230
Manual ability				
	1	0.000		
	2	-0.035	-0.055	-0.016
	3	-0.112	-0.132	-0.092
	4	-0.226	-0.245	-0.207
	5	-0.294	-0.315	-0.272
Communication				
	1	0.000		
	2	-0.021	-0.041	-0.002
	3	-0.063	-0.081	-0.045
	4	-0.181	-0.201	-0.162
	5	-0.218	-0.238	-0.198
Pain and Discomfort				
	1	0.000		
	2	-0.040	-0.060	-0.020
	3	-0.065	-0.085	-0.044

	4	-0.097	-0.117	-0.077
	5	-0.301	-0.324	-0.277
Sleep				
	1	0.000		
	2	-0.028	-0.047	-0.008
	3	-0.092	-0.112	-0.073
	4	-0.190	-0.210	-0.170
	5	-0.253	-0.274	-0.232

Figure 5.1 shows the Australian utility algorithm for the conditional logit (mixed logit utility decrements are in the appendix which took into account the heterogeneity of the references).
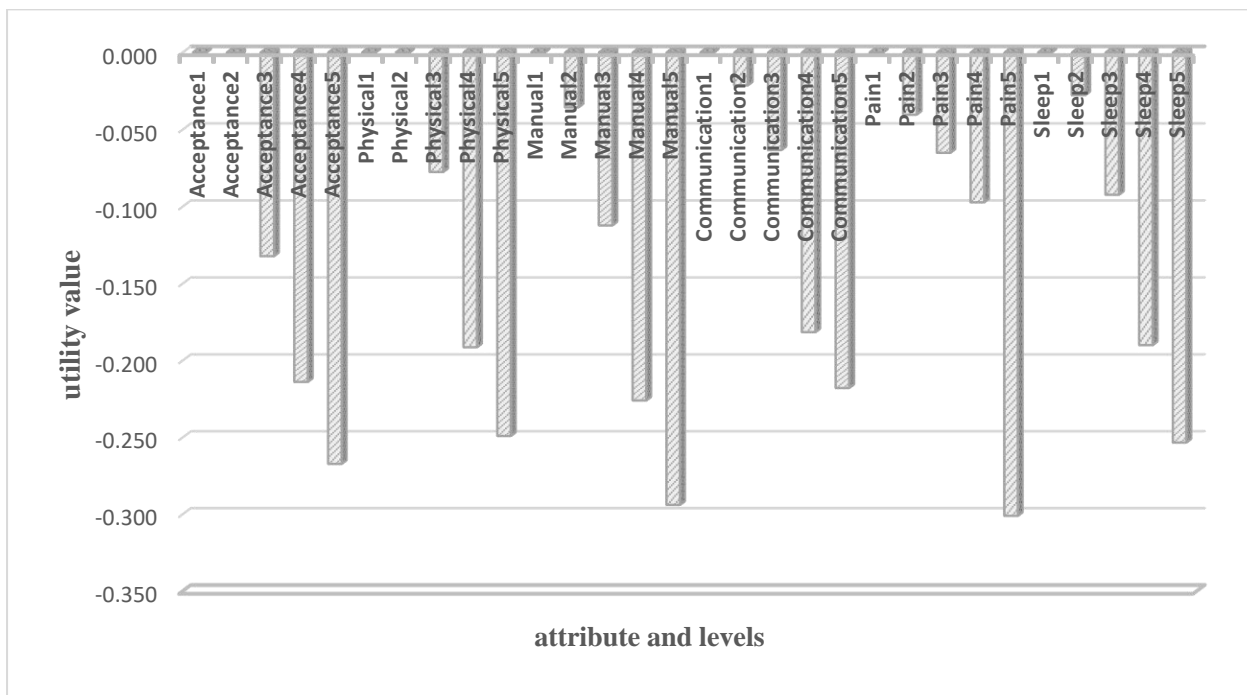
Figure 5.1 Australian Utility Algorithm for CP-6D derived from conditional logit

CP-6D utility calculation

As mentioned previously, the utility value for full health (level 1 of each dimension) was fixed at one. The utility for other health states would be equal to one minus the sum of the utility decrements of the CP-6D domains levels. The utility value for health state 152241 based

on the calculation method in the results would be $1 - (0 + 0.249 + 0.035 + 0.021 + 0.097 + 0) = 0.598$. The value of the worst possible health state, 555555, is -0.582.

The results for the lexicographic preferences demonstrated that the alternative-specific constant for health state A or health state B was not statistically significant (p-value > 0.05); ~1.5% of participants always chose the same side health state (23 people always selected health state A; seven always chose health state B). All participants chose the lowest duration (shorter life expectancy) at least once. About 28.5% of respondents failed the consistency check.

We did not exclude any participants from the study. However, we ran the model several times by removing the duration-based responses, removing the people who did not answer the repeated question similarly separately. However, the characteristics of the model did not change.

Response time and difficulty

The average time spent on each choice set is presented in Figure 5.2; the highest average belongs to the first choice set and it decreases as participants progress through the tasks. The highest and lowest average times are 55.5 seconds and 21.3 seconds, belonging to first and thirteenth choice sets, respectively. Most people found both answering and understanding the questions easy (49% and 35%); 3.5% found the questions difficult to understand and 5% found them difficult to answer.
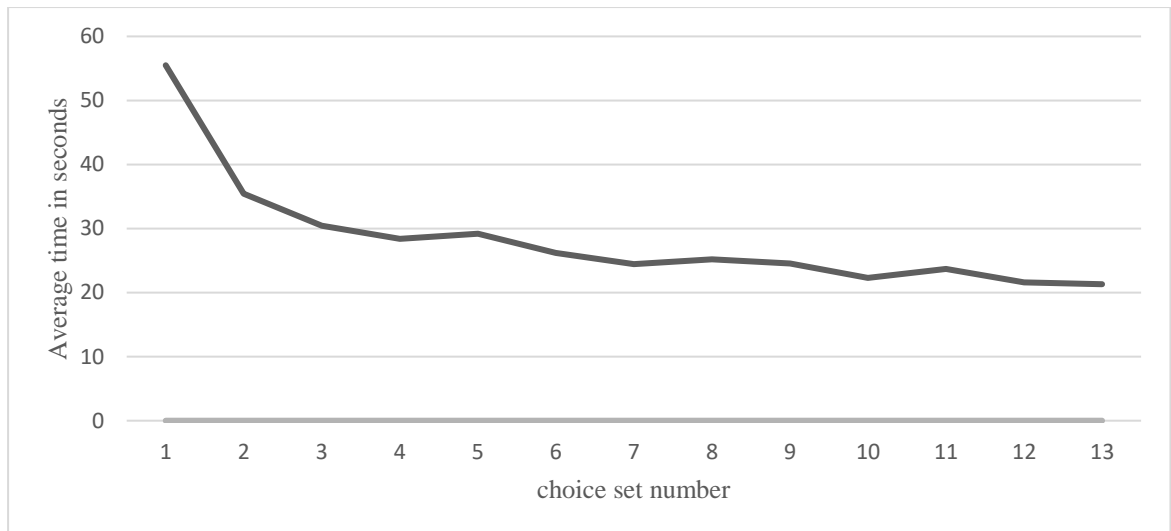
Figure 5.2 Average time spent on each task in the DCE valuation tasks

5.5 Discussion

The results of this study present the first utility value set for the CP-6D, a multi-attribute utility instrument derived from the CPQOL. To develop the CP-6D, the CPQOL proxy version was used with the parent projection and happiness regarding the child's health state being measured. The resulting values can potentially be applied in economic evaluation studies. A condition-specific utility instrument can capture the most important attributes related to that condition or disease and assign them weights based on their importance. However, as outlined in the introduction, there is an ongoing debate about using condition-specific versus generic instruments. The utility scores generated from condition-specific instruments might be more sensitive to changes in QoL caused by that condition (183), as they capture the condition-specific attributes; these estimates can be used for scenario analysis but at the potential cost of generalisability of findings (38, 39).

The severity of CP can be classified by the Gross Motor Functioning (GMFCS) (227), Manual Ability (MACS) (205) and Communication Function (CFCS) (252) instruments. All three were adapted as attributes in the classification system. CP QoL cannot be directly

measured by the widely used instruments, such as the EQ-5D, as it does not have communication, sleep or manual ability, which are important domains (the usual activities and self-care domain might capture some aspects of it) (253), especially in people with CP (215). The HUI3—the most-used instrument for QoL in children with CP—has a dexterity domain comparable with manual ability (254) but does not include sleep or communication domains.

An unusual feature of the CPQOL is that response levels are based on happiness, unlike other QoL instruments where response levels are based on case severity. Happiness has been used interchangeably with well-being and QoL as a factor of well-being (255). Using happiness as a response level is thought to capture a person's true well-being and utility (256). Therefore, happiness levels might capture a broader aspect of well-being and not just health alone. Further, using happiness as a response level combines both physical and mental health in one aspect; thus, applying it as a level might measure a broader component of HRQoL. Conversely, it is likely that happiness with a particular aspect of HRQoL will not perfectly correlate with severity. A study reported that people who adapt to their health status have healthier and longer lives, and experience a significantly higher lifetime utility than non-adapting types (257); if adaption can modify health utility, then being happy about one's health might have the same effect. Thus, exploring how well happiness levels capture the HRQoL is an avenue for future studies.

The utility value set range is wider here than is observed in other instruments; the worst health state is valued at -0.582. For comparison, the worst health state in the Canadian HUI3 valuation study is -0.36 (254) and the value for the EQ-5D-3L for Australian population using DCE is -0.516 (245). This difference in utility values and pits state can be caused by a range of factors(258), such as the study location, the valuation method, the attributes in the descriptive system, and the sample population. The larger range in the utility scale, which is

shown by a lower pits state, indicates the greater average marginal change in utility score per change in health state, which can highlight differences when conducting an economic evaluation based on CUA (112).

The manual ability domain had the largest utility decrements in this study. Manual ability is a CP-sensitive domain and its inclusion in the descriptive system provides a more relevant utility measure for CP. It refers to a person's physical dexterity and requires fine hand motions. This functioning is crucial for many daily activities (259) and problems due to hand motions can affect a person's QoL.

Acceptance is another domain likely to be important in a CP context. In people with disabilities like CP, potential acceptance issues can start at early ages, and children with CP may feel excluded and unaccepted in some circumstances at home or school. At home, they might be unable to interact with their siblings in the same way as others and in school, they might need special classes or devices that their peers without disabilities do not need. Therefore, feeling accepted can make them happy and increase their HRQoL (260).

Sleep is a dimension that exists in some generic MAUIs and is likely an important dimension in people with CP. Sleeping problems are more frequent in children with CP; this might be due to physical (e.g., other comorbidities) or environmental factors (bedding or equipment) (261). Physical health is presented by how a person feels about getting around and is comparable with the HUI3 ambulation domain or EQ-5D mobility domain. It is how the child feels about moving around and applies to both children who can walk and those who use special equipment, such as crutches and wheelchairs, which are classified by the GMFCS.

DCEs generally assume that respondents compare attributes and levels and choose their preferred option from a set of health states; however, some studies report that participants

employ lexicographic decision-making rules and that decisions are biased (149). This may not be correct. For example, choosing the longer duration can be a genuine choice and the respondent may actually prefer that health state over the alternative. However, in this questionnaire, there are no wrong and right answers. Consequently, even though respondents answered the repeated question differently, we did not delete them from the analysis.

This study is not without limitations. Like other DCE studies, respondent behaviour was not fully evaluated while completing the survey and we cannot fully comprehend the extent to which they understood the questions. In addition, Lim et al (262) suggest that efficient design algorithms in the context of DCEtto studies of a specific format (triple sets with full health as an option) favour inclusion of milder states over severe states and so, the range of severity is not fully covered; this can be a disadvantage of DCEtto designs and can be checked in further study when designing a DCEtto.

The study respondents were a large sample representative of the Australian population in age and sex; hence, the non-representativeness of their other demographic data (e.g., ethnicity or location) may be a disadvantage that could be explored in future work. Use of the general population as our sample may also be a drawback because there is an ongoing debate on who should value health for economic evaluation (263). However, in Australia (39), the standard base case is to use the value set from a general population in cost-utility analyses intended to inform societal decisions. However, we acknowledge that a lived experience of the condition can inform people about the severity of different QoL aspects and hence, may generate different valuations. Future valuation studies could be performed using participants who are familiar with the condition to examine whether the values differ.

One important consideration is whether the CP-6D can be administered as a stand-alone instrument or whether its questions must be considered in the context of the broader CPQOL.

Stand-alone administration would be preferable in terms of respondent cognitive burden; however, the questions asked immediately before items from the CP-6D might affect how respondents perceive those items. Therefore, we suggest future research explore the comparability of data collected using both approaches (260). It could be also beneficial to apply a psychometric comparison between CP-6D with generic measures for further assessment of this instrument's performance.

5.6   Conclusion

This paper contains the CP-6D value set for an Australian population based on their preferences. The results are applicable in economic evaluations for CP interventions to determine value for money. They will be useful for Health Technology Assessment and other health care agencies during decision-making processes regarding resource allocation for people with CP. Consequently, models to predict outcomes and care costs for CP will help inform health provision and social care, while providing data for national funding schemes.

6 Chapter 6: Comparing multi-attribute utility instruments: CP-6D vs AQoL-4D

STATEMENT OF CONTRIBUTION TO CO-AUTHORED PUBLISHED PAPER

This chapter includes a published co-authored paper. The bibliographic details of the co-authored paper, including all authors, are:

Bahrampour M, Downes M, Scuffham PA, Byrnes J. Comparing multi-attribute utility instruments: CP-6D vs AQoL-4D.

My contribution to the paper involved the designing the study and the survey, planning and executing the data collection, analysing the data and drafting the article.

(Signed) _____ (Date)_____

Mina Bahrampour

(Countersigned) _____ (Date)_____

Corresponding author of paper: Mina Bahrampour

(Countersigned) _____ (Date)_____

Supervisor: Prof. Paul Scuffham

Comparing multi-attribute utility instruments: CP-6D vs AQoL-4D

6.1 Abstract

Background: Economic-evaluations of Cerebral palsy (CP) were based on utility estimates of health-related quality of life (HRQoL) from generic multi-attribute utility instruments (MAUIs). However, generic instruments had limited use as they could not capture some of the important aspects of living with CP. The Cerebral palsy 6 Dimension (CP-6D) is a condition specific MAUI. In this study, we compared the results of CP-6D with the Assessment of Quality of Life (AQoL-4D), a generic MAUI, and tested the criterion validity of the CP-6D in the general population.

Methods: An online survey of the Australian general population (n=2002), who completed both the AQoL-4D and CP-6D MAUIs, was conducted. Validity was assessed from the correlations between the domains, items and instruments. ANOVA and t-tests were used to assess the instrument's discrimination in different social demographic categories.

Results: There was a moderate correlation between the instruments (0.64). Differences in socio-demographic characteristics showed a medium effect size ($p < 0.001$) in both instruments and had a similar effect on utility weights in both instruments. Although, the CP-6D was more sensitive to changes in income and education.

Conclusions: Our results suggest that CP-6D and AQoL-4D were measuring a similar underlying construct. Both instruments responded similarly to socio-demographic differences.

Keywords: Utility, Validity, CP-6D, AQoL-4D, Health related quality of life, Cerebral palsy, Preference based measure

6.2 Introduction

Generic preference based measures (PBMs) or multi-attribute utility instruments (MAUI) are useful in valuing HRQoL and for use in economic evaluations within the context of health technology assessment (31, 264). In particular, generic instruments are designed to measure, value and compare health states in the general population and across a wide range of health conditions (41, 265). However, in recent years, condition-specific PBMs have been developed due to generic measures' lack of specificity in certain conditions. Condition-specific instruments take into account the precise domains of HRQoL that are more relevant to the particular condition and therefore are expected to be more sensitive to changes in the latent construct of HRQoL than generic instruments, within that sub-group of the population (266).

CP affects an estimated 0.2% live births worldwide annually and is the most common disability in children (2). CP is an umbrella term for a group of disorders caused by a lesion in the developing brain that affects the movement of the body and, in some cases, causes intellectual impairment. The Cerebral Palsy 6 Dimension (CP-6D) is a new condition-specific measure of HRQoL in people with CP. It has six domains; each domain contains one item and each domain has five response levels (215). The CP-6D instrument values health states according to a preference-based utility index that ranges from one to zero (representing full health and death respectively), but also with negative values that represent states worse than death (51, 52).

The AQoL-4D is a widely used generic MAUI (267) which describes HRQoL in four domains: independent living, relationships, mental health, and physical senses (hearing, vision and speech). Each domain contains three items, and each item has four response levels (268). The AQoL-4D also values health states according to a preference-based utility index that ranges from one to less than zero (51, 52).

Validity is about how well an instrument measures the areas of interest, meaning to what extent an instrument measures what it is supposed to measure (269, 270). Validity is often thought of as a multi-dimensional construct with validity of an instrument considered with respect to content, construct and criterion. Content validity refers to the items and domains of an instrument with respect to their suitability and whether an instrument is qualified for what it is intended to measure, which is usually done using experts' opinions (including patients as experts). Construct related validity seeks the agreement between a theoretical concept and a specific measure, with discrimination being the recognition of the differences between various categories. The larger the effect size, the better the discriminating ability (271). Criterion validity is the extent to which a measure is related to an outcome. As part of this, concurrent validity evaluates the strength of the relationship between instruments that have the same concept (measured using Pearson correlation). Where correlations between instruments indicate that the PBMs are assessing related constructs.

Over the past decades, to compare MAUIs, the main focus has been on the validity and sensitivity of the instruments (272-274); although more recently, performance related to relative discrimination has been used to compare the performance of different instruments (275). During development the CP-6D has been validated implicitly using experts' opinion (content related validity), factor analysis and Rasch analysis (construct related validity). The reliability of the instrument was also evaluated using Cronbach's alpha (~0.89).

However, before widely applying this new condition-specific instrument, we need to further test its construct validity and criterion validity. To date, the measurement properties of the CP-6D have not been compared to those of a generic PBM; therefore, we sought to compare the utility values estimated using the CP-6D with the AQoL-4D utility values and test the criterion validity of the

CP-6D in a general population. The AQoL-4D was used as it was one of the four preference based instruments that have been used in the CP population (56).

6.3 Methods

An online survey of both the AQoL-4D and the CP-6D along with demographic factors such as age, gender, income level, level of education, and having private health insurance was conducted during September and October 2019. This study was conducted in a sample of 2002 members of the Australian general population above the age of 18 years old. The sample was selected using a random quota procedure, based on the distribution of the population-specific characteristics of gender and age within the Australian population based on the Australian Bureau of Statistics (ABS). The online survey administration was conducted by SurveyEngine, a company with expertise in online surveys.

AQoL-4D: The AQoL-4D items were derived de novo from reviewing existing HRQoL instruments (276), and these fit with the definition of health by the World Health Organization (WHO) (267). The AQoL-4D has 12 items that are categorised into four domains: independent living, relations, mental health and the senses. Self-care, household tasks and mobility comprise the independent living domain; friendships, isolation and family comprise the relations domain; sleeping, worrying and pain comprise the mental health domain; and, seeing hearing and communication comprise the senses domain. The value set for the AQoL-4D in the Australian general population ranges from -0.04 to 1 (276).

CP-6D: The CP-6D has been developed from the Cerebral Palsy Quality of Life Questionnaire (CPQOL), a CP specific instrument, using factor and Rasch analysis. The CP-6D has six domains, each represented by one item. These domains are: social wellbeing and acceptance, physical

health, manual ability, communication, pain and discomfort, and sleep (215). To estimate utility values for this instrument, a discrete choice experiment was used. The scores of the CP-6D range between -0.582 and 1 in the Australian general population (277).

Statistical analysis: Correlations between the CP-6D and the AQoL-4D utility values, dimensions, and items were evaluated independently. If correlation scores are less than 0.3 the correlation is considered weak, if scores are between 0.3 and less than 0.7 the correlation is moderate, and scores of 0.7 or higher indicate a strong correlation (278). The differences and discrimination observed in the instrument's scores among each income class and education group were tested using ANOVA. The effect size was used to explore the size of the observed difference of summary scores between the classes. Eta squared (η^2), which represents the proportion of the variance in the dependent variable, is an estimate of effect size when analyzing the variance in (ANOVA) models. A larger effect size signifies better discriminating ability; the effect size is categorised into small (0.01), medium (0.05) and large (0.14) (279). For private health insurance, a t-test was applied, and Cohen's d showed the effect size. Cohen's d works better with larger samples and shows the difference in the means in the variables. Using Cohen's d, a rule of thumb for effect size is that less than 0.2 is considered a small effect, 0.5 is a medium effect and 0.8 is a large effect (278).

Age, income and education were all categorized into six groups, while gender and having private health insurance were binary choices. Annual income was categorized as: less than $38,000; $38,000 to $70,000; $70,000 to $112,000; $112,000 to $205,000, and more than $205,000, we also had a "prefer not to say" option. Education contained these classes: Year 11 or below, Year 12, trade certificate, diploma, bachelor's degree, and postgraduate degree. Age was categorized into: 18 to 29, 30 to 39, 40 to 49, 50 to 59, 60 to 69, and 70 and above.

6.4   Results

The online survey included 2002 individual participants. The sample represents the general

Australian population in age and gender; about 51 percent of the participants were women, and

the mean age was ~46 years. The demographics of the sample are provided in Table 6.1.

Table 6.1 Sample characteristics

Variables		Study population (n)	Study population (%)
Gender	Male	976	48.75
	Female	1026	51.25
Age (years)	18-29	419	20.93
	30-39	362	18.10
	40-49	347	17.33
	50-59	327	16.33
	60-69	273	13.63
	70 and above	274	13.68
Education	Year 11 or below	250	12.49
	Year 12	354	17.68
	Trade certificate	287	14.33
	Diploma	319	15.93
	Bachelor's degree	539	26.92
	Postgraduate degree	253	12.65
Income	Less than $38,000	428	21.37
	$38,000 to $70,000	537	26.82
	$70,000 to $112,000	450	22.48
	$112,000 to $205,000	345	17.23
	More than $205,000	94	4.70
	refer not to say	148	7.40
Private health insurance	Yes	1169	58.39
	No	833	41.61

For the full sample, the mean AQoL-4D utility score was 0.635 (SD=0.280) and 0.721 (SD=0.265)

was the mean CP-6D utility score. The correlation between the two instruments was direct and

positive with r=0.637 (p <0.001). This correlation demonstrates that both instruments are

measuring a similar latent variable in the same direction. The spread of the utility values of both

instruments is presented in Figure 6.1. In both instruments, ~50% or more of the population had a

utility value of more than 0.6.
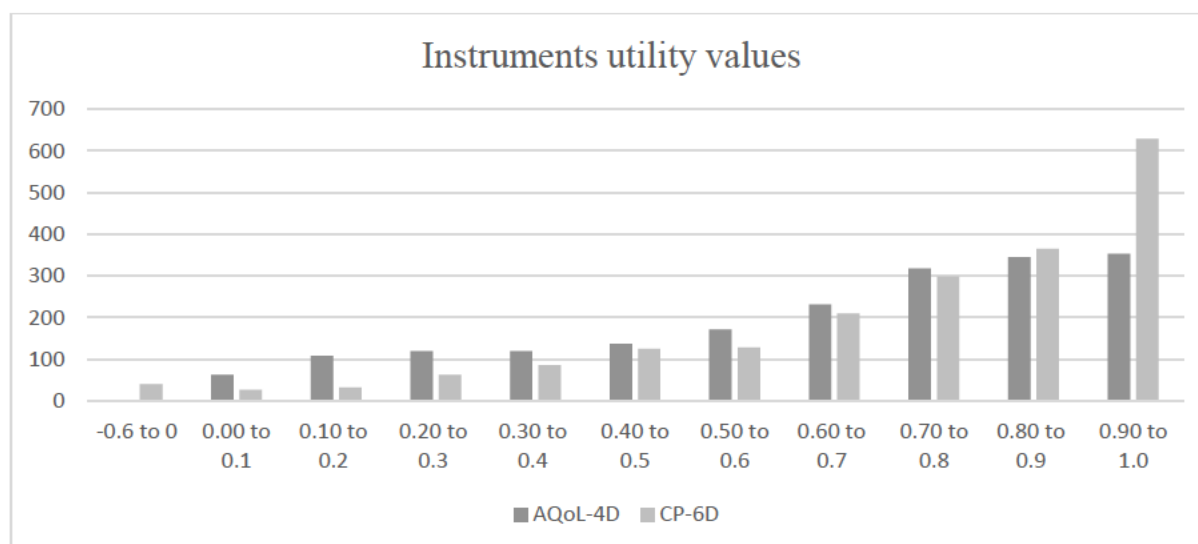
Figure 6.1 Utility values of both AQoL-4D and CP-6D

Agreement between instruments

The correlation between all the domains was positive (p <0.001). The correlations between the CP-6D and the AQoL-4D dimensions presented in Table 2 indicates a low level of overlap except for the mental health dimension. The AQoL-4D's mental health had the highest correlation with the CP-6D's domains of *pain and discomfort* and *sleep* respectively. The CP-6D's *social wellbeing and acceptance* and *communication* domains both had the highest correlation with the AQoL-4D's relations domain.

Table 6.2 Correlation between domains

AQoL domains \ CP-6D domains	Independent living	Relationships	Senses	Mental health
Social wellbeing and acceptance	0.119	0.448	0.183	0.311
Physical health	0.335	0.382	0.209	0.406
Manual ability	0.326	0.302	0.237	0.306
Communication	0.210	0.417	0.289	0.290
Pain and discomfort	0.410	0.337	0.235	0.551
Sleep	0.142	0.301	0.131	0.441

p < 0.001 for all coefficients of correlation
CP-6D:Cerebral palsy 6 dimension
AQoL-4D: Assessment of Quality of Life 4 dimension

Table 6.3 shows the correlation coefficient between the six items of the CP-6D and the 12 items of the AQoL-4D, which are mostly low (below 0.4). However, the items of *sleep* and *pain* in both instruments correlated moderately, with the coefficients of 0.676 and 0.575 respectively. *Communication* items in the two instruments had a weak correlation.

Table 6.3 Correlation between items

AQoL / CP-6D	Acceptance by people	Physical health	Manual ability	Communication	Pain	Sleep
looking after yourself	0.045*	0.198	0.233	0.115	0.295	0.066*
Household tasks	0.094	0.341	0.334	0.189	0.408	0.168
Get around	0.096	0.257	0.270	0.176	0.327	0.096
Relationship	0.392	0.335	0.273	0.388	0.290	0.262
Other relations	0.440	0.355	0.251	0.400	0.316	0.315
Family relation	0.256	0.378	0.350	0.294	0.395	0.222
Vision	0.197	0.187	0.150	0.187	0.167	0.155
Hearing	0.086	0.143	0.159	0.153	0.154	0.100
Communication	0.146	0.176	0.243	0.276	0.222	0.068*
Sleep	0.247	0.316	0.239	0.183	0.312	0.676
Emotion	0.418	0.370	0.209	0.350	0.340	0.402
Pain	0.157	0.388	0.310	0.165	0.575	0.303

*p-value > 0.001

Discrimination

An analysis of the sensitivity of each instrument to changes in the subgroups of income, education, and having private health insurance is presented in Table 6.4. Both instruments showed that they are sensitive for private health insurance, income, and education. The effect size for all variables was higher in the CP-6D. Both instruments recorded a higher quality of life with increasing income and education (Table 6.4). The change of utility in both instruments was statistically significant ($p < 0.001$). The table indicates that income had the highest effect size in both instruments.

Table 6.4 Discrimination across socio-demographic variables

	Definition	CP-6D		AQoL-4D	
		Mean (SE)	Effect size	Mean (SE)	Effect size
Private health insurance	No	0.685(0.010)	0.23	0.600(0.090)	0.21
	Yes	0.746(0.007)		0.659(0.008)	
Income	Less than $38,000	0.615(0.015)	0.05	0.537(0.013)	0.04
	$38,000 to $70,000	0.716(0.011)		0.634(0.011)	
	$70,000 to $112,000	0.764(0.010)		0.652(0.013)	
	$112,000 to $205,000	0.783(0.010)		0.714(0.012)	
	More than $205,000	0.817(0.019)		0.682(0.034)	
	Prefer not to say	0.704(0.022)		0.651(0.021)	
Education	Year 11 or below	0.565(0.0184)	0.03	0.633(0.0203)	0.01
	Year 12	0.639(0.0146)		0.702(0.0143)	
	Trade certificate	0.629(0.0164)		0.695(0.0167)	
	Diploma	0.650(0.0150)		0.724(0.0156)	
	Bachelor's degree	0.650(0.0117)		0.757(0.009)	
	Postgraduate degree	0.650(0.0185)		0.778(0.0131)	

The difference in the quality of life utility value is equivalent between the two instruments; however, the absolute value differs. Increasing quality of life is mostly associated with higher income and advanced education (Figures 6.2 and 6.3).
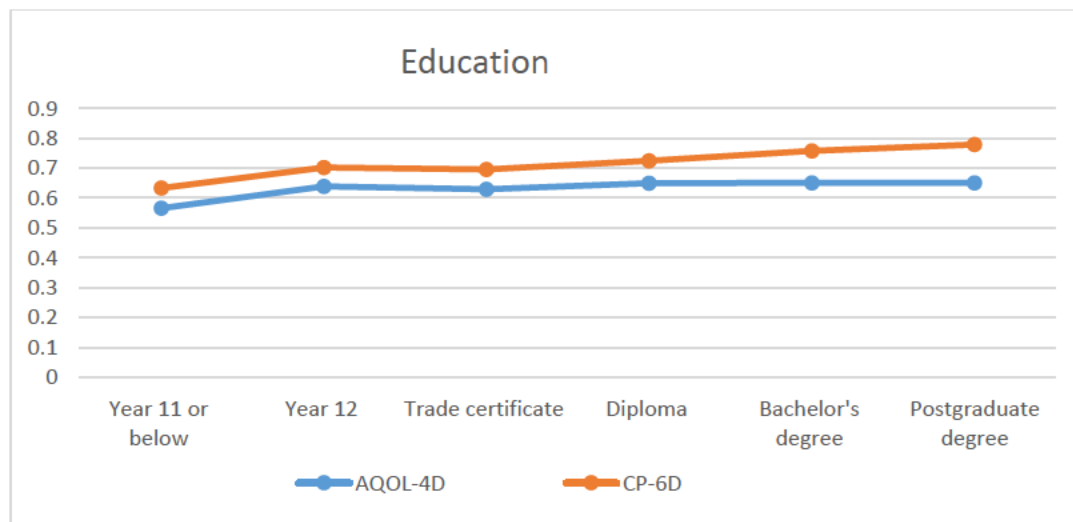
Figure 6.2 Changes in utility value different education groups
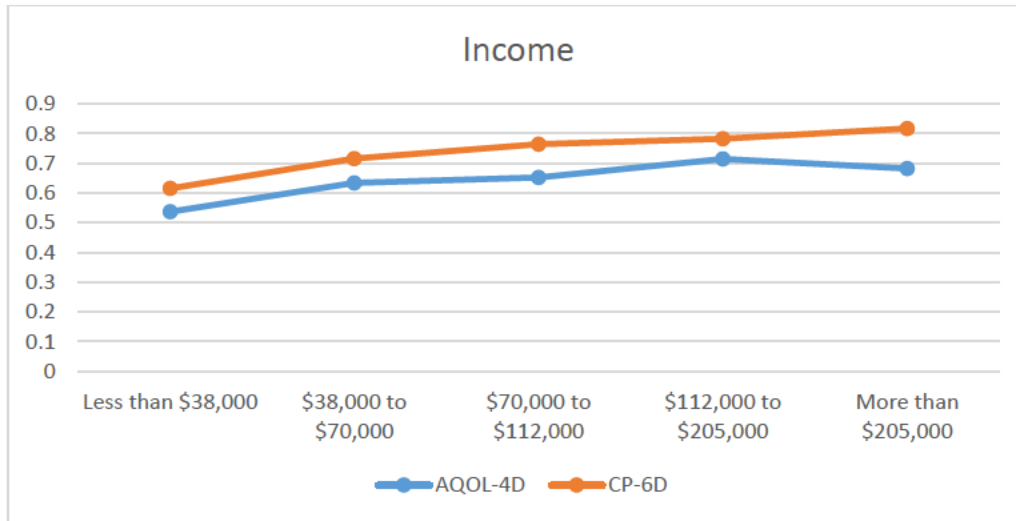
Figure 6.3 Utility value trend by income

Both instruments recorded a higher quality of life when people had private insurance compared to no private insurance (Figure 6.4).
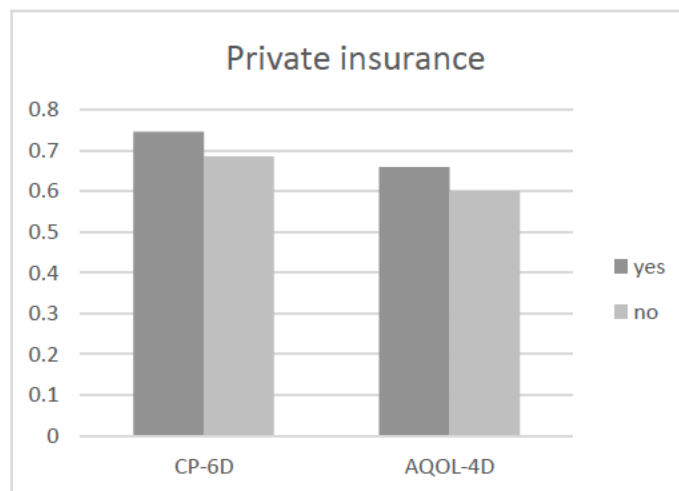
Figure 6.4 Effect of private insurance on utility

6.5   Discussion

The purpose of this study was to compare the construct validity and criterion validity of two instruments. In this study, for the first time, the variation of utility values estimated by the AQoL-

4D and the CP-6D in a general population setting have been tested and compared. The key finding is that the mean utility value in both instruments were similar and that there is reasonable construct validity between AQoL-4D and CP-6D. Higher income and education, and having private health insurance had a positive effect on the utility weights from both instruments.

As for criterion validity, the low coefficient of correlation between items and between domains can be due to the structure of the instruments as the domains are not exactly the same. For instance, in the CP-6D the items *pain* and *sleep* are items for two separate domains, whereas, in AQoL-4D, these two are sub-items under the same domain, which is *mental health*. There are also a range of sub-items within the AQoL-4D's various domains which do not map across to the CP-6D. Another difference between the instruments is the response levels. In the CP-6D, response levels are reported on a happiness scale (very happy = 1 to very unhappy = 5), whereas the AQoL-4D levels are based on a severity or frequency scale. When looking at the correlations between the *emotion* item of AQoL-4D and CP-6D items, the emotional item is correlated among all the items of the CP-6D, which would indicate that the happiness scale used in the CP6D may be moderated by the responders emotion. However, the estimated correlations were low and therefore does not appear to be a significant confounder in the responses to the CP-6D.

Considering the difference in discrimination between the two measures, attention needs to be given to the exact wording of the instruments. For instance, even though both instruments have the item *communication,* the wording of the questions is very different. In the CP-6D, individuals are asked about how they feel about how people communicate with them, whereas in the AQoL-4D, the participants are asked about how they communicate with people. Also, the recall period, meaning the time interval which participants are asked to reflect upon when responding to an item, differs between the instruments. The CP-6D asks the participants to think about their current health,

115

whereas the AQoL-4D instructs them to consider the past week. This difference in perspective is likely to influence how people feel about and value their health (280).

As the utility values are at the higher ends of the spectrum in the CP-6D (see Appendix), this raises a concern for a ceiling effect within the scores. Despite the greater spread of the AQoL-4D participants across the utility scale, the CP-6D is more sensitive, as evidenced by the greater effect size when it comes to changes in socio-demographic characteristics. The effect size is slightly higher in the CP-6D, which may mean that the CP-6D is more sensitive than the AQoL-4D with respect to changes in socio-demographic categories.

There are some similarities and differences between the instruments, which impacts the psychometric features, and consequently may affect cost-effectiveness analyses of technologies to improve HRQoL. The construct of HRQoL in the AQoL-4D is different from that in the CP-6D. Among its items, the AQoL-4D includes vision and hearing. Regardless of the variation in domains and items, and the small correlation between the items and between the domains, the instruments' utility weights are moderately and directly correlated. This has also been shown in other studies and suggests that even though the instruments are measuring a similar concept there are distinct differences, which would be expected (274). Furthermore, it has been shown that various MAUIs are not equally responsive to changes (56) and that different measures give various utility values (276). This also affects the result of CUA and the funding of services.

This study was conducted in the general population rather than the people who had the condition in the case of this study CP. Often, as in previous studies, the comparison between a condition-specific instrument and a generic instrument are done in the patient population (275, 281, 282), and the results used to assess the validity and sensitivity of the instrument in different severity groups (53). Researchers have sought to measure the improvement in sensitivity within the

condition group of interest and have confirmed that condition-specific instruments will lead to different utility estimates within the condition population compared to generic instruments. They have also illustrated that the condition-specific instrument is more sensitive than the generic instrument in the condition groups (60).

The majority of generic instruments have been assessed in the general populations, however, to our knowledge, no studies have attempted to compare a condition-specific instrument to a generic instrument in a general population, which can be comprised of mixed conditions; so as to assess the extent that the condition-specific instruments are able to capture domains of general importance in defining and measuring HRQoL(60). Hence using the general population to understand construct validity before applying the instrument to a patient group is an appropriate method because it demonstrates the appropriateness of the instruments (283, 284). Using the general population to validate condition specific instruments can be an avenue for future studies when constructing a validity study.

Of note, the mean value of AQoL-4D in this study showed a lower value compared to the Australian population norms using the AQoL-4D (285); however, these results are similar to the results of a more recent online survey by McCallum et al. [16], which suggests they are consistent with the online Australian population.

A limitation of this study can be the absence of the CP population to measure the utility values in this group. However, this aspect of current study may provide an avenue for future research which is how well the CP-6D instrument measures utility in a CP population. Using the general population gives confidence about the overall and general performance of the instrument; yet, more research is needed to understand how this instrument captures utility in the CP population. This study provides supporting evidence that a condition-specific instrument may be comparable

with a generic instrument. The differences between the instruments demonstrates that the CP-6D can be more sensitive than the AQoL-4D, however more studies can be done to check the sensitivity of the CP-6D in different severity groups.

6.6   Conclusion

The results confirm the correlation between the CP-6D and the AQoL-4D suggesting that they are measuring a similar underlying construct. This study adds further weight to the validity of the CP-6D. Based on discriminatory and correlation analysis, both construct and criterion-related validity are confirmed illustrating that the CP-6D can be used as an HRQOL instrument.

7   Discussion

7.1   Introduction

In this chapter the most important aspects of this study to develop a CP-specific preference based utility measure are brought together. To reach the overall aim of this thesis, several questions have been introduced and answered in each of the preceding chapters. A detailed discussion of the specific findings and their implications for each research question have been provided in the chapters which have been presented as journal articles. In this chapter, the findings are reviewed, summarized and evaluated to determine the extent to which the study's aims have been met. This chapter also includes the strengths and limitations of the study, a conclusion and suggestions for future research.

7.2   Background

7.2.1   Preference based measures in cerebral palsy research

Cerebral palsy (CP) is the main cause of childhood disability with a prevalence of 2.1 per 1000 live births (173). CP may occur during pregnancy or shortly after birth (18) and is a costly condition (20). The health related quality of life (HRQoL) of individuals with CP is mostly affected by motor functioning issues (286). This condition is permanent, and it has extensive impacts on all aspects of an individual's daily life as well as the lives of their families (287, 288). From a broader perspective, the effect of CP is also seen in the education and welfare systems. Given the burden and impact of CP, more improvement is crucial to guide interventions and clinical therapies for this condition. Hence, it is of principal importance to ensure that the effectiveness of interventions and treatments and their expected costs are underpinned by research that is reliable and valid for use by clinicians, researchers and policymakers.

Economic evaluation helps to compare and find the most cost-effective interventions and clinical therapies. The use of health economic evaluation to assist decision-making has increased (289). One of the main methods of economic evaluation is cost-utility analysis (CUA), which uses quality-adjusted life years (QALYs) to measure health outcomes when allocating resources. CUA is the preferred technique by most health technology assessment (HTA) agencies around the world to decide about health care interventions (290-293). QALYs measure the health outcomes by combining the quality of life with the length of life. The benefit of using QALYs is that they allow comparison of cost-effectiveness across health conditions, which can help with resource allocation and health care planning (294).

The quality of life (QoL) can be measured by utility values. Utility values are generated using preference weights that inform scoring algorithms. The weights are ranged on a scale of zero to one, where zero indicates death and one is full health; states worse than death can be represented, these states take a negative value. Preferences for health states are obtained using a preference elicitation method such as time trade-off (TTO), standard gamble (SG) and discrete choice experiment (DCE). Utility values are obtained by preference based measures. A preference based measure consists of two components: a classification system to define health states and a set of utility weights to estimate utility values for the health state (60, 295, 296).

Generally, the preference based measures that have been used in CUA are generic, and the most used one is the EQ-5D. When EQ-5D is found to be unsuitable for a condition or patient group, alternative methods are recommended. Using an instrument that does not cover essential aspects of a condition will lack sensitivity in the measurement. It might result in inaccurate utility values, potentially overestimated values, that will affect the estimated QALYs that will be used

in CUA (60). Alternatives to the EQ-5D are other generic preference based measures or condition-specific preference based measures. There has been an increase in using preference based condition-specific instruments and methods to develop these instruments (60). However, there are still some concerns about the validity and responsiveness of using condition-specific measures to compare interventions for health conditions (242, 297).

It has been determined that five studies identified the utility values associated with CP in children (57). In these studies, direct health state valuation methods or generic preference based measures were used to estimate QALYs. Of these, four have used generic measures (58, 173, 174, 298), and one used a direct method to measure values (299). There was no condition-specific preference based measure available for CP. It has been illustrated that among the preference based PROMs to estimate utility values for CP, which were all generic instruments, the Health Utility Index Version 3 (HUI3) was the most promising instrument for children with CP (56). However, it had some disadvantages in this population; for instance, it had a poor performance in individuals who used wheelchairs (300). In previous studies, it was mentioned that the commonly used preference based measures did not capture all important HRQoL domains for children with CP and there is a lack of information demonstrating the instruments' responsiveness and reliability (56, 57). As the previously existing tools were not capable of capturing some of the important domains that impacted the HRQoL of people with CP, a sensitive and specific instrument to measure HRQoL in CP was needed for CUA of interventions for the CP population.

A comprehensive systematic literature review was conducted to identify the multi-attribute utility instruments that used DCEs as a method to generate preference based values, with the aim of finding the best method to derive a DCE (Chapter 2). The systematic review results

demonstrated that DCE method has not been used to value utilities for CP when using a CP-specific preference based measure.

Further, to the best of our knowledge, there were no condition-specific preference based measures (CSPBM) available for people with CP. Also, to promote the economic evaluation of treatment and interventions using CUA among people with CP, there is the necessity of utility measures to quantify outcomes in terms of QALYs in the CP population. The main aim of this thesis was to fill this gap by developing a CP-specific preference based quality of life measure to be used in CUA of the interventions for CP. The first step to a preference based measure is to develop a CP-specific classification system which is what has been done in this thesis.

7.3   Development of a preference based quality of life measure

7.3.1   Development of a cerebral palsy classification system for a preference based quality of life measure

A CP-specific health state classification system was developed to generate health states amenable to valuation. The classification system, called the CP-6D, was developed from the CPQOL using Rasch and factor analysis (Chapter 3). The first step to develop utility weights for a certain condition is to employ a classification (descriptive) system (54, 301, 302) that contains the most important aspects related to that condition. The condition-specific preference based instrument can be developed de novo or from an existing QoL instrument.

As there was a reliable and valid non-preference based, pediatric QoL measure in children with CP, called the CPQOL-child (parent/proxy version) (167), this instrument was used to develop the classification system. The CPQOL has been developed to assess the wellbeing of

children with CP using the views of parents /proxies and children (64). This instrument has been used in CP studies and provides a QoL score between zero and 100. The main reason for the CPQOL development was to understand whether an intervention has changed a child's QoL or not. However, all the items of the CPQOL are equally weighted (62). For instance, the improvement in the child's feeling about their pain level has the same weighting as the improvement in the child's feeling when they miss school. This makes it challenging to distinguish the importance of different interventions. As the attributes do not have the same effect on a person's HRQoL, the current system of scoring in the instrument is insufficiently reflective of what is important. Hence, it does not indicate the potential magnitude of what people are willing to sacrifice to move from one health state to another. Utility weights are essential to compare between interventions in a CUA (168). Because the CPQOL is not able to provide utility weights, it cannot be used to estimate QALYs. Therefore, a preference based measure is required to assess QALYs to evaluate the effectiveness of interventions in a CUA framework (303).

A classification system was developed using cohort CPQOL data by applying factor and Rasch analysis (Chapter 3), which would be used to subsequently derive utility weights for each health state described by the classification system. The advantage of using an existing instrument to develop a classification system is that the scoring algorithm and utility values from the generated instrument can be used for existing studies that have applied the existing instrument (60). Furthermore, the classification system is the shorter form of the original instrument and can be used in clinical trials, which will be less time consuming and it can be applied to estimate the QALYs for economic evaluation. Rasch and factor analysis have been applied in previous studies to develop a classification system using existing QoL instruments and have showed reliable performance (176, 301).

The CPQOL has 65 items in seven domains: social wellbeing and acceptance, participation and physical health, feelings about functioning, emotional wellbeing and self-esteem, pain and impact of disability, access to services, and family health. By applying factor and Rasch analysis and expert opinion, a new instrument was generated that is called the Cerebral Palsy Six Dimension (CP-6D). The CP-6D has six domains, some of which differ from the original instrument. Four domains of the CP-6D were similar to the CPQOL, which were derived as a result of the factor and Rasch analysis and two extra domains (manual ability and sleep) were added based on expert opinion because of their importance for children with CP. The domains of the CP-6D were thus social wellbeing and acceptance, physical health, manual ability, communication, pain and discomfort, and sleep. All the steps in the development ensured the face, content and construct validity of the instrument (Chapter 3).

The CPQOL used a 9-point Likert scale, but the descriptors were provided for only five of the levels (for instance Level 1 was very happy, but Levels 2 and 3 both showed as happy, so the disordering happened between Levels 2 and 3), which made it confusing for participants and caused item level disorder after Rasch analysis. Disordering in response levels can happen when participants have difficulty in distinguishing between levels (176). Due to disordering in the levels after the Rasch analysis, through consensus among the expert panel and the Rasch analysis, the nine-level response scale was reduced by combining some of the disordered levels, and a five-level classification system was adopted. The response levels of the CP-6D were very unhappy, unhappy, neither unhappy nor happy, happy and very happy. This was with the exception of the domain of pain and discomfort, in which the levels were very upset, moderately upset, somewhat upset, slightly upset and not at all upset. The exact wording of the CPQOL items and response

levels were used, and the wording of the original instrument was not altered during development (62).

Clinically, the severity of CP can be classified by the Gross Motor Functioning (GMFCS) (12), Manual Ability (MACS) (13) and Communication Function (CFCS) (15) classification systems. All three were adapted as attributes in the CP-6D. Manual ability has been added as a domain in the new instrument, as using hands is a method of classifying the severity of the CP in children. Communication was another new domain; however, the items regarding communication were previously a priori subsumed in the CPQOL domain of feelings about functioning. For the GMFCS the domain of physical health was used, which is a domain from the original instrument. Sleep was another new domain; this domain was added as children with CP have more problems with sleep than their peers (304).

There is an emphasis on emotional wellbeing and feelings when developing instruments for children with CP (64, 305). Frequently, the main focus of QoL is physical functioning (306). Happiness, however, is a broader concept of QoL, and it comprises the satisfaction of doing things in addition to the ability to do them (307). Using happiness as a response level captures the person's true level of wellbeing (256), as happiness shows wellbeing. Studies have demonstrated a high correlation between the happiness and health domains of the QoL (308). The CP-6D levels that are based on happiness could capture a broader aspect of wellbeing and not just health alone. Happiness as a response level was considered to combine physical and mental health in one aspect; therefore, mental health was not added as a separate domain in the CP-6D.

The final instrument consists of six domains, and each domain was represented by one item which had five response levels. A CP-specific classification system is the first step of developing

a CP-specific preference based measure that can be valued using preference elicitation methods. The next step is performing a valuation study to develop a utility set for the new instrument and generate QALYs that can be used in the economic evaluation of interventions for people with CP.

7.3.1.1 Preference elicitation methods

Utility values can be produced using different preference elicitation methods. Different elicitation techniques can be cardinal, such as the time trade-off (TTO) or standard gamble (SG). In TTO, the tasks require the participant to tradeoff between life years, and SG assesses respondents' preference for health by determining the amount of risk of "immediate death" they are willing to accept. Or they can be ordinal methods such as a discrete choice experiment (DCE) and best worst scaling (BWS). In DCEs, the participants are asked to choose their preferred health state among scenarios, and in the BWS method, the respondents are asked to select the best and worst attributes of a health state (104). As the TTO tasks require participants to trade life years, and SG has the risk of immediate death as an option, conducting a valuation study for the paediatric population will be challenging. Adolescents have found the BWS and DCE task easier to understand and complete (66). In recent years there has been an increasing interest in employing DCE over other methods due to its easy administration using an online platform (79).

Usually, in DCEs the participants are asked to choose their preferred health state between hypothetical scenarios. This preference based method is based on attributes. It can measure benefit (utility), and can also be used for different objectives such as prioritizing healthcare based on attributes of care delivery (309). When valuing health states using DCEs, there can be different methods; hence some issues need to be considered, such as DCE designs, the number of choice sets, the analyzing method and the anchoring technique.

126

```
┌─────────────────────┐      ┌─────────────────────┐      ┌─────────────────────┐
│  Identifying relevant│ ───> │ Assigning levels to  │ ───> │ Applying statistical │
│     dimensions       │      │   these dimensions   │      │    design theory     │
└─────────────────────┘      └─────────────────────┘      └─────────────────────┘
                                                                      │
        ┌─────────────────────┐      ┌─────────────────────┐         │
        │ Presenting choice    │ ───> │ asking respondents to│ <───────┘
        │ sets to respondents  │      │ express their        │
        │                      │      │ preferences by       │
        │                      │      │ choosing within the  │
        │                      │      │ choice sets.         │
        └─────────────────────┘      └─────────────────────┘
```
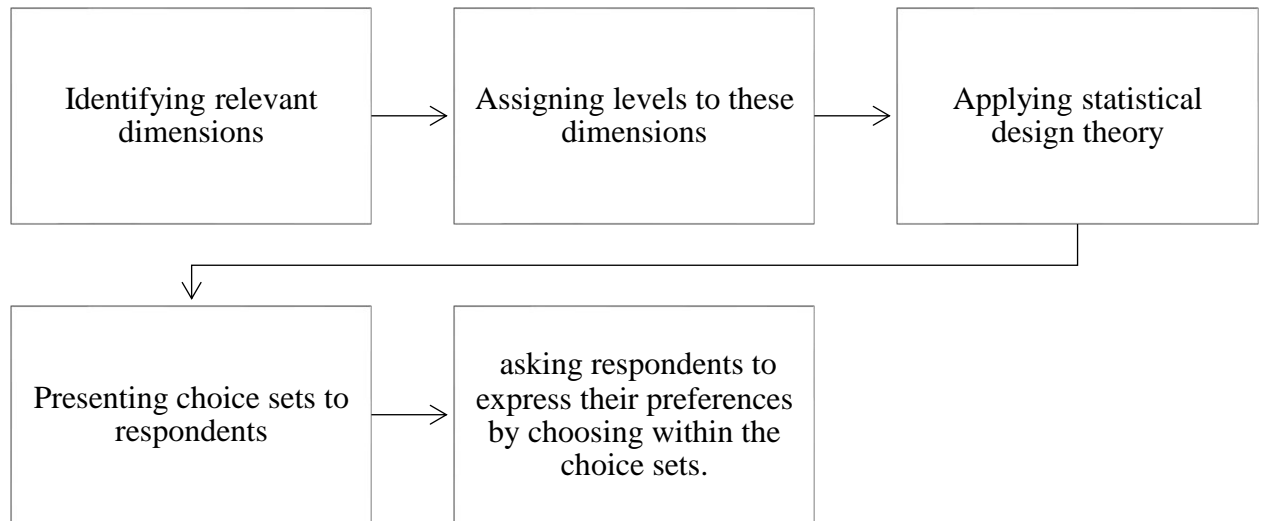
Figure 7.1 Framework for applying DCEs

7.3.1.2   Application of discrete choice experiments when valuing multi-attribute utility instruments

A systematic review was conducted to find the best method of applying a DCE to develop utility values for a preference based measure, also known as a multi-attribute utility instrument (MAUI) (Chapter 2). In the study, even though no one gold standard method was demonstrated, the most frequent methods were mentioned, and the advantages and disadvantages of each technique were illustrated. In this review it was found that overall, 38 studies had used DCE to estimate utility values in MAUIs, based on the studies' inclusion and exclusion criteria.

Different techniques have been used to measure utility using DCEs. While there is no consensus on the ultimate method for this goal, there were similarities among studies as well as differences. Similarly to another review (79), it was found that the majority of studies reported using an online panel and a sample of the general population. Differences were in terms of how to design the choice sets, how to anchor the values, how to check consistency, how many choice sets

are possible, how to minimize the number of choices without a lot of bias and how to analyze the data.

The included studies (38 studies) consisted of generic and condition-specific instruments; the number of the condition-specific instruments, however, was low. The EQ-5D has been identified as the most common MAUI (79), which concurs with the findings of the systematic review (chapter two). Some studies reported that EQ-5D did not involve all the main domains of a specific condition, thus condition-specific instruments for some conditions, such as cancer (302, 310) and asthma (176), were developed.

A debate is going on about who should value health states to estimate the utility values. On one hand, some researchers argue that the general population's opinions matter as they are the tax-payers and pay the funds. On the other hand, it is debated that the person who has the condition or is familiar with the condition (parent/ proxy/ caregiver) should be the one valuing the health states. However, it may be that not one nor the other option is best; but rather, both are relevant but should be used for different purposes. This debate is still ongoing for valuation studies of MAUIs.

Conditional logit was the most used method to analyse DCE studies; this method has been mostly used in choice behaviour studies (82). Conditional logit takes the society's interest into account, and economic evaluation holds a greater interest in the mean response and the differences among the choices of the society (98). There were different methods to anchor the utility weights from the DCE on a 0 to 1 scale (104, 246); the DCEtto method was the most used approach to anchor the utility values. In this method, the coefficients of the conditional logit were estimated when the duration was added as an attribute when designing the study. Even though DCEs were used in valuation studies to overcome other methodological limitations, using them alone for

valuing health has its analytical challenges, especially when anchoring (79, 153). As DCEs produce utility values on a latent scale, the utility values obtained must be anchored on a full health and dead scale prior to estimating QALYs (311)

Considerable methodology and design diversity existed in the literature. The differences in the methodology and design of the DCEs will be considered useful by some researchers. It will promote learning about the impact of methodological choices and help guide decisions about future protocols or elements of a new methodology that may apply to a wide range of DCE studies. Because of the diversity of methods, users need to understand the features and limitations of value sets before using them in decision-making. Researchers need to continue to strive for value sets that are valid and reflect true preferences while continuing to educate and be guided by decision-makers and the public.

Despite there being no "best method" recommendation, the increasing number of publications in which DCEs were used for valuing health states shows that DCEs are potentially valuable for developing value sets and will continue to be used as a method to value health state utilities. While the variation in methods remains, more development is required.

7.3.2 Valuation of a preference based quality of life measure for the cerebral palsy six dimension (CP-6D)

For the valuation study, the DCE method with duration (DCEtto), was administered to an Australian general population using an online survey to estimate the utility values (Chapter 5). DCEs have been used to value health in several studies in recent years (121, 138, 245, 312). For the sake of simplicity, online administered DCEs were used instead of TTO and SG methods to value health states. Online data collection substantially reduces the time and resource required

when compared to other approaches such as face-to-face interviews and postal or telephone methods (313). Duration was added as an attribute in the DCE design (DCEtto) to help anchor the values to measure QALYs. The study was conducted in two steps.

The first step was a pilot to measure the priors for the main and final design for estimating the utility values. The sample in the pilot study were 80 individuals above 18 years old from the Australian general population. The pilot was used to check the difficulty of the questions, to indicate the feasibility of the duration levels, to assess the time spent on the survey, and to determine the functioning of the survey.

The second part was the final data collection from the Australian general population through a web-based DCE study. DCEs are based on a simple and straightforward task; so compared to other elicitation methods, they are easier to administer (69, 79). Hence, a web-based method was suitable for the DCE survey (150); furthermore, participant recruitment was easier online.

In Australia (39), in CUA, the standard base case is to use the value set from a general population to inform decisions. Hence, the data from the general population were used to ensure that the preferences of voters and tax-payers were reflected in the utility weights obtained. It goes without saying that the characteristics of the population used to produce health state utility values could affect the results of the CUA and thus potentially affect resource allocation in decision-making (54). As indicated elsewhere, participants over 18 were recruited, as the choice tasks may be methodologically challenging for children, and children have limited cognitive ability to complete the valuation tasks. Even many adolescents have indicated that utility valuation tasks were challenging to complete (314).

A sample of 2002 people from Australian general population were selected randomly in this study. However, a quota was added in the online survey to ensure the sample represented Australia in age and gender. Based on suggestions in the literature, the sample size needed to be large enough to ensure there were enough respondents per choice set (112, 234).

Conditional logit was used to model the DCE data and estimated coefficients. Dimensions were generally monotonic, meaning that the coefficients reflected the ordered and expected direction nature of the levels in each dimension, in that more severe levels had greater utility decrements. All coefficients were statistically significant, except the result for Level 2 of physical health, which was not statistically significant.

The utility decrement from full health for each level of each domain was estimated. The higher level illustrates more decrement in the utility value from full health; in this study, the best level was very happy (=1). The decrement from full health indicates that each level of each attribute can be responsible for the change in the utility value. The best health state or full health was shown as 111111, and the worst level was 555555. The utility weights for the CP-6D health states ranged from -0.582 to 1.

The manual ability dimension was associated with the highest utility decrements compared to other dimensions. In a recent study in which a mapping algorithm was developed from the CPQOL onto the CHU-9D (315), a children generic preference based measure, it was reported that the upper limb could be more important for HRQoL than gross motor function in the CP population. This adds weight to the current study's finding for manual ability.

There were no previous studies that valued health states derived from a classification system in CP, so a direct comparison of the utility values obtained from this study with previous research could not be done.

7.3.2.1 The psychometric performance of the CP-6D and the AQoL-4D in the general population:

In Chapter 6, the CP-6D was compared to a generic preference based measure (the AQoL-4D) for validation. The validity of an instrument is the extent to which that instrument measures what it is designed to measure. Usually, to evaluate the validity in patient-reported outcomes, methods are derived from the theories and techniques of psychological measurement. It was decided to compare the psychometric performance of the CP-6D with a generic preference based instrument, the AQoL-4D, as the early evidence of validity and responsiveness of this condition-specific preference based measure.

Methods to understand validity differ. Validity can be content-related, which shows that the items and domains of an instrument are suitable and accurate, and the instrument is qualified for what it is intended to measure. Face-related validity is similar to content validity, but it is more informal. There is not a specific statistical measure to evaluate both content and face validity and it is usually performed by asking experts. Criterion-related validity evaluates the extent to which the result of one measurement is related to another (criterion) instrument; to evaluate criterion validity the correlation between two instruments can be calculated. Another method is construct-related validity in which agreement is sought between a theoretical concept and a specific measure.

The internal consistency of the CP-6D with CPQOL in terms of face and content validity was ensured using factor and Rasch analysis and expert opinions when developing the

classification system (215) (Chapter 3). After developing the utility algorithm for the CP-6D, the utility values for the general population were measured. The utility values for the AQoL-4D for this population was also estimated using the weighted scoring algorithm (276).

To evaluate the construct validity of the instrument, the correlation between the CP-6D and AQoL-4D, their domains, and their items was estimated. The results showed a moderate and direct correlation between the instruments, but the correlation between items and between the domains was low. The low correlation between items and domains has been shown in other studies (275, 281, 282), and in this case it can be readily explained. The low correlation shows that although the instruments are measuring a similar variable, the aspects of HRQoL may vary. This also happened when comparing two generic instruments (316). The changes in utility values were similar when the categories of socio-demographic variables changed. In this study, when income and education were higher, the average utility weights generated from both instruments rose, which is consistent with other studies (317).

The variations between the wording of the items limits any comparative analysis of surveys (318). Even the slightest change in the wording of a question can result in differences in the responses obtained (319). Therefore, the different wording of domains or items presenting a domain can change the meaning completely and may cause varying utility values, even for the same person. For instance, both AQoL-4D and CP-6D have the domain of communication, however, the item wording is different. In CP-6D, individuals are asked about how they "feel about how people communicate with them", while in AQoL-4D the participants are asked about how they communicate with people.

CP-6D and AQoL-4D have different recall times, which might be one of the reasons the correlation is not high. The recall time of the instrument can affect how a person values their health and the health state the participant reports. It was demonstrated that an instrument with a longer recall period gave lower utility values (280). If the recall period is more than one day it requires the participant to report their health status over a time which might include a fluctuation in their health and cause recall bias. When considering the period of illness in retrospect, participants tend to remember more negative experiences and extreme health states compared to more stable levels (320, 321). In sum, there are a range of explanations for variance in the psychometric performance of the two instruments.

7.4   Policy implications

This study fills a gap in the literature and indicates the need for comprehensive information for better health policy decision-making for the CP population. This study provides the first CP-specific classification system for children with CP. The CP-6D survey, which is a short version of the CPQOL instrument, can be time-saving when measuring the quality of life in children with CP. The CP-6D can be used in preference based measurement and generate QALYs for children with CP. In addition, the outcomes of the valuation study can be transferable both prospectively and retrospectively to CPQOL data. This will help to inform future planning and resource allocation for CP care. Moreover, the information gained can also improve the decision-making process and offering evidence-based policy implications.

In Australia, in 2018, the total financial cost of cerebral palsy was estimated to be $3.03 billion, with a further $2.15 billion in lost wellbeing (21). Significant costs include lost productivity ($1.29 billion), costs of disability support services ($1.12 billion) and efficiency

losses (\$553.59 million). This CP-specific instrument and utility values would assist in identifying the most cost-effective interventions, treatments and clinical therapies. Hence this instrument would support the decision-making process to allocate resources effectively within a limited budget. Hence, the CP-6D will facilitate policy decisions that will have a positive impact on improving the lives of individuals with CP and their families.

7.5   Strength and limitations

In this thesis a comprehensive systematic review was conducted of DCE methods used to value health states in preference based measures and a detailed overview was given of the methods used for the relevant instruments. Moreover, the studies' reporting quality was assessed. The current research is the first study to develop a CP-specific preference based measure (classification/descriptive system). Also, it is the first study in which DCE was used to generate the utility values for CP using the preferences of the general Australian population, which can be incorporated into economic evaluations for the intervention and treatment for people with CP. The psychometric performance of the CP-specific instrument was also compared with a generic preference based instrument, and differences in preferences were highlighted among socioeconomic groups according to income and education.

As with any research, there were some limitations when accomplishing this study. The first is related to the lack of an external data set to validate the CP-6D classification systems. When deriving a condition-specific preference based measure from an existing instrument, this is generally an important step; but, it depends on the data availability to researchers. In this innovative study, no external CPQOL data were available to validate the classification system. An alternative in this case is the split-half approach, in which half of the data are used to develop the

classification system and the other half is used to validate it. The low number of data restricted the use of a split-half approach to validate the instrument. The ideal sample size for Rasch analysis is ~ 400-500, but in this study, splitting the data into two would have resulted in unsuitably small data sets for Rasch analysis. However, triangulation through the factor and Rasch studies and expert opinion ensured face validity and supported the contention that the most important items have been included in the classification system.

It was mentioned in the protocol (Chapter 4) for the valuation study that the patient group utility values would also be estimated and compared with the utilities generated from the general population. Although the survey was sent to three Australian states' CP registries, a very low number of surveys were completed (<20). Unfortunately, because of insufficient data this aim was not met and the valuation has not been done by neither proxies nor the children with CP. In the questionnaire designed for this population we asked the parent that if their child has the ability to reply and response to questions, ask them to participate with the parent, using a parent-child dyad. This method was designed to involve the child in the discussion while they represent their responses.

Another limitation of this work is that the validity of the CP-6D was only tested in the general population. As validity is a test of whether CP-6D measured the intended construct, it would have been ideal to conduct this test in the target population (children with CP or the parents). Future research should consider replicating this analysis in a CP population.

7.6   Conclusion and suggestions for future studies

The use of DCEs to value health states has increased in recent years. A comprehensive systematic review was conducted to find the best method to use DCEs when valuing utilities in a

preference based measure. The literature review conducted in this research provided a detailed summary of the methodological features of DCEs for the valuation of health states. It was demonstrated that despite the variety of methods, the overall framework is the same. For all studies, the users need to understand what they want to measure so they can choose what kind of design is required, what type of anchoring will suit their aim, how to check logical consistency, and what the best method is to analyse the data. Also, in this study, it was shown that there was no CP-specific instrument that used DCE to value health.

This study is the first to develop a preference based measure, the CP-6D, for the CP population using a classification system that was generated by applying factor and Rasch analysis along with consultation with patients and clinicians and seeking expert opinions. The classification system consists of six items, with each item having five response levels. As an output of the research, the first CP-specific utility value set was also provided. DCE was used to estimate the Australian-specific utility weights of CP-6D; the utility values ranged from -0.582 to 1.000. These values can potentially be used in CUA of interventions for people with CP among the Australian population.

The CP-6D was valued by the adult general population, however, a valuation study using participants who are familiar with the condition and comparing the values of both groups can be a promising avenue for future research. In children's surveys, mostly a proxy is asked to complete the questionnaire as the survey questions are usually difficult to respond to. However, future studies can estimate the values using both proxy and children preferences, the results of the two group can be compared, which can be beneficial to distinguish the differences in both population groups.

In this study, the validity of the CP-6D was also assessed. The CP-6D content-related validity was evaluated at the development stage by using expert opinions. Its construct-related validity was assessed by using a generic preference based measure, the AQoL-4D, in a general population. A correlation was demonstrated between the instruments for construct validity.

The CP-6D offers potential advantages over generic measures and we recommend the CP-6D to be used together with generic measures in future studies. Using the CP-6D alongside generic measures ensures that the needs of policymakers can be met, but also offers an excellent opportunity to explore the comparability of data collected and the utility values generated from both instruments. This will also provide further evidence for testing this new instrument and examining its impact on CUA.

In the future, researchers could use a range of indicators to investigate the discriminative ability of the CP-6D utility measures as there is no clear outline for it at the time of writing. Furthermore, additional research needs to be performed on the discriminatory ability of the instrument among different severity groups, as there was insufficient data generated to draw any conclusions for these groups.

Consequently, government bodies such as the Medical Services Advisory Committee (MSAC), Pharmaceutical Benefits Advisory Committee (PBAC), and pharmaceutical companies can use the utility values estimates when they are applying a CUA for the CP population as a part of an HTA. The utility weights can be applied in an economic evaluation of new treatments or interventions for children with CP; these weights show the change in a persons' health state when a new treatment or intervention is applied. The utility values can be used in CUA for CP

interventions where the CPQOL or the CP-6D is applied. This helps to provide economic evidence

to allocate resources for interventions and treatments for people with CP.

In estimating the utility values for the CP-6D, the preferences of the Australian general

population have been considered. Generalizability to other countries is not yet demonstrated.

Therefore, a cross-culture validation should be performed to generate the values for other regions,

which will result in deriving utility algorithms for other countries. Such algorithms can estimate

values reflective of the population the decision-makers serve.

It is recommended that those who are designing clinical trials and studies on interventions

for CP consider the use of the CP-6D alongside the CPQOL. CP-6D usage to measure HRQoL for

different intervention for children with CP will help to assess the robustness of this new instrument,

for example, in terms of reliability, predictive, convergent and discriminant validity. In sum, to

clarify the benefits and weaknesses of the new measures developed and to check the performance

and acceptability of the new instrument, further research is encouraged.

References

1.	Rosenbaum P, Paneth N, Leviton A, Goldstein M, Bax M, Damiano D, et al. A report: the definition and classification of cerebral palsy April 2006. Developmental Medicine & Child Neurology Suppl. 2007;109(suppl 109):8-14.
2.	Reddihough D. Cerebral palsy in childhood. Australian Family Physician. 2011;40(4):192-6.
3.	Reddihough DS, Collins KJ. The epidemiology and causes of cerebral palsy. Australian Journal of Physiotherapy. 2003;49(1):7-12.
4.	Nelson KB. Causative factors in cerebral palsy. Clinical Obstetrics and Gynecology. 2008;51(4):749-62.
5.	McIntyre S, Morgan C, Walker K, Novak I. Cerebral palsy--don't delay. Developmental Disabilities Research reviews. 2011;17(2):114-29.
6.	Jacobsson B, Hagberg G. Antenatal risk factors for cerebral palsy. Best Practice & Research Clinical Obstetrics & Gynaecology. 2004;18(3):425-36.
7.	Sankar C, Mundkur N. Cerebral palsy-definition, classification, etiology and early diagnosis. Indian Journal of Pediatrics. 2005;72(10):865-8.
8.	Pakula AT, Braun KVN, Yeargin-Allsopp M. Cerebral palsy: classification and epidemiology. Physical Medicine and Rehabilitation Clinics. 2009;20(3):425-52.
9.	Taylor F. Cerebral palsy: hope through research. National Institute of Neurological Disorders and Stroke (US), Office of Science and Health Reports.Bethesda, Md.: The Institute, 2001. Accessed online September 28, 2005; 2006.
10.	Hutton JL, Pharoah POD. Life expectancy in severe cerebral palsy. Archives of Disease in Childhood. 2006;91(3):254-8.
11.	Holmes Jr L, Joshi A, Lorenz Z, Miller F, Dabney K. Pediatric cerebral palsy life expectancy: has survival improved over time. Pediat Therapeut. 2013;3(1):1-6.
12.	Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. Developmental Medicine & Child Neurology. 1997;39(4):214-23.
13.	Eliasson A-C, Krumlinde-Sundholm L, Rösblad B, Beckung E, Arner M, Öhrvall A-M, et al. The Manual Ability Classification System (MACS) for children with cerebral palsy: scale development and evidence of validity and reliability. Developmental Medicine & Child Neurology. 2006;48(7):549-54.
14.	Morris C, Kurinczuk JJ, Fitzpatrick R, Rosenbaum PL. Reliability of the manual ability classification system for children with cerebral palsy. Developmental Medicine & Child Neurology. 2006;48(12):950-3.
15.	Hidecker MJC, Paneth N, Rosenbaum PL, Kent RD, Lillie J, Eulenberg JB, et al. Developing and validating the Communication Function Classification System for individuals with cerebral palsy. Developmental Medicine & Child Neurology. 2011;53(8):704-10.
16.	Russo RN, Goodwin EJ, Miller MD, Haan EA, Connell TM, Crotty M. Self-esteem, self-concept, and quality of life in children with hemiplegic cerebral palsy. The Journal of Pediatrics. 2008;153(4):473-7.
17.	Livingston MH, Rosenbaum PL, Russell DJ, Palisano RJ. Quality of life among adolescents with cerebral palsy: what does the literature tell us? Developmental Medicine & Child Neurology. 2007;49(3):225-31.
18.	Oskoui M, Coutinho F, Dykeman J, Jette N, Pringsheim T. An update on the prevalence of cerebral palsy: a systematic review and meta-analysis. Developmental Medicine & Child Neurology. 2013;55(6):509-19.
19.	The Australian Cerebral Palsy Register Group. The Australian Cerebral Palsy Register Group. Australian cerebral palsy register report 2016: Birth years Australia: The Australian Cerebral Palsy Register (ACPR) Group; 2016.p.7.

20.   Cerebral Palsy Australia. The Economic impact of cerebral palsy in Australia in 2007. Access Economics Pty Ltd; 2008. p. 74.
21.   The cost of cerebral palsy in Australia in 2018. Deloitte Access Economics; 2019.
22.   Pareezer L, Jordan, R., Ware, R., Boyd, R.N. Cost of medical and allied health resource use and consequence in preschool age children with cerebral palsy. Developmental Medicine and Child Neurology. 2012;54:71-2.
23.   Jordan R, Finn, C., Arnfield, E., Fiori, S., Guzzetta, A., Ware, R., Boyd, R.N. Relationship between health resource use and performance outcomes in preschool-aged children with cerebral palsy: An economic analysis. Developmental Medicine and Child Neurology. 2014;56:43-4.
24.   Honeycutt AA, Grosse SD, Dunlap LJ, Schendel DE, Chen H, Brann E, et al. Economic costs of mental retardation, cerebral palsy, hearing loss, and vision impairment. Using survey data to study disability: results from the national health survey on disability. Emerald Group Publishing Limited; 2003. p. 207-28.
25.   Economic costs associated with mental retardation, cerebral palsy, hearing loss, and vision impairment--United States, 2003. MMWR Morbidity and Mortality Weekly Report. 2004;53(3):57-9.
26.   Kancherla V, Amendah DD, Grosse SD, Yeargin-Allsopp M, Van Naarden Braun K. Medical expenditures attributable to cerebral palsy and intellectual disability among Medicaid-enrolled children. Research in developmental disabilities. 2012;33(3):832-40.
27.   Banta D. The development of health technology assessment. Health Policy. 2003;63(2):121-32.
28.   Richardson J, Schlander M. Health technology assessment (HTA) and economic evaluation: efficiency or fairness first. J Mark Access Health Policy. 2018;7(1):1557981-.
29.   Scuffham PA, Whitty JA, Mitchell A, Viney R. The use of QALY weights for QALY calculations. Pharmacoeconomics. 2008;26(4):297-310.
30.   Dolan P, Shaw R, Tsuchiya A, Williams A. QALY maximisation and people's preferences: a methodological review of the literature. Health Economics. 2005;14(2):197-208.
31.   Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes: Oxford University Press; 2015.
32.   Dakin H, Wordsworth S. Cost-minimisation analysis versus cost-effectiveness analysis, revisited. Health Economics. 2013;22(1):22-34.
33.   Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG. Cost-effectiveness in health and medicine: Oxford University Press; 2016.
34.   National Institute for Health and Care Excellence. Appendix G: Methodology checklist: economic evaluations (PMG6B). 2012.
35.   Harris AHB, Liliana. Getting value for money: the Australian experience. In: Jost TS, editor. Health care coverage determinations: an international comparative study. Berkshire UK: Open University Press; 2005. p. 25-56.
36.   Utility[online]. York: York Health Economics Consortium; 2016 [Available from: http://www.yhec.co.uk/glossary/utility/.
37.   Cubbon J. The principle of QALY maximisation as the basis for allocating health care resources. J Med Ethics. 1991;17(4):181-4.
38.   National Institute for Health and Care Excellence (NICE). Technical guidance for manufacturers and sponsors onmaking a submission to a technology appraisal. London: NICE;2009.
39.   Commonwealth of Australia. Guidelines for the pharmaceutical industry on preparation of submission to the pharmaceuticals benefits advisory committee:including submission involving economic analysis. Canberra: Australian Government Printing Office;2016.
40.   Guidelines for the economic evaluation of health technologies: Canada(Ottawa): 4th ed. Canadian Agency for Drugs and Technologies in Health (CADTH) 2017.
41.   Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. Measuring and valuing health benefits for economic evaluation.2nd ed: Oxford University Press; 2017.

42.   Prieto L, Sacristán JA. Problems and solutions in calculating quality-adjusted life years (QALYs). Health & Quality Of Life Outcomes. 2003;1:80.

43.   Sinnott PL, Joyce VR, Barnett PG. Preference Measurement in Economic Analysis. Guidebook. Menlo Park CA. VA Palo Alto: Health Economics Resource Center; 2007.

44.   What quality of life? The WHOQOL Group. World Health Organization Quality of Life Assessment. World Health Forum. 1996;17(4):354-6.

45.   Ferrans CE. Definitions and conceptual models of quality of life. Outcomes assessment in cancer: Measures, methods, and applications. 2005:14-30.

46.   Romero M, Vivas-Consuelo D, Alvis-Guzman N. Is Health Related Quality of Life (HRQoL) a valid indicator for health systems evaluation? SpringerPlus. 2013;2(1):664.

47.   Rockville M. Patient-reported outcomes, health-related quality of life, and function: an overview of measurement properties In: Feeny DH, Eckstrom E, Whitlock EP, Perdue LA, editors. A primer for systematic reviewers on the measurement of functional status and health-related quality of life in older adults. US: Agency for Healthcare Research and Quality; 2013.

48.   Lin X-J, Lin I-M, Fan S-Y. Methodological issues in measuring health-related quality of life. Tzu Chi Medical Journal. 2013;25(1):8-12.

49.   Kind P, Lafata JE, Matuszewski K, Raisch D. The Use of QALYs in Clinical and Patient Decision-Making: Issues and Prospects. Value in Health. 2009;12(s1).

50.   Sassi F. Calculating QALYs, comparing QALY and DALY calculations. Health Policy and Planning. 2006;21(5):402-8.

51.   Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. Br Med Bull. 2010;96(1):5-21.

52.   Neumann PJ, Goldie SJ, Weinstein MC. Preference-Based Measures in Economic Evaluation in Health Care. Annual Review of Public Health. 2000;21(1):587-611.

53.   Lorgelly PK, Doble B, Rowen D, Brazier J. Condition-specific or generic preference-based measures in oncology? A comparison of the EORTC-8D and the EQ-5D-3L. Quality of life research. 2017;26(5):1163-76.

54.   Mulhern B, Rowen D, Brazier J, Smith S, Romeo R, Tait R, et al. Development of DEMQOL-U and DEMQOL-PROXY-U: generation of preference-based indices from DEMQOL and DEMQOL-PROXY for use in economic evaluation. Health Technology Assessment. 2013;17(5).

55.   Stevens KJ. How well do the generic multi-attribute utility instruments incorporate patient and public views into their descriptive systems? The Patient - Patient-Centered Outcomes Research. 2016;9(1):5-13.

56.   Mpundu-Kaambwa C, Chen G, Huynh E, Russo R, Ratcliffe J. A review of preference-based measures for the assessment of quality of life in children and adolescents with cerebral palsy. Quality of Life Research. 2018. doi:10.1007/s11136-018-1837-0.

57.   Tonmukayakul U, Le LK-D, Mudiyanselage SB, Engel L, Bucholc J, Mulhern B, et al. A systematic review of utility values in children with cerebral palsy. Quality of Life Research. 2018;28(1):1-12.

58.   de Lissovoy G, Matza LS, Green H, Werner M, Edgar T. Cost-effectiveness of intrathecal baclofen therapy for the treatment of severe spasticity associated with cerebral palsy. Journal of Child Neurology. 2007;22(1):49-59.

59.   Brazier J, Czoski-Murray C, Roberts J, Brown M, Symonds T, Kelleher C. Estimation of a preference-based index from a condition-specific measure: the King's Health Questionnaire. Medical Decision Making. 2008;28(1):113-26.

60.   Brazier J, Rowen D, Mavranezouli I, Tsuchiya A, Young T, Yang Y, et al. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). Health technology assessment. 2012;16(32):1-114.

61.   Carlon S, Shields N, Yong K, Gilmore R, Sakzewski L, Boyd R. A systematic review of the psychometric properties of Quality of Life measures for school aged children with cerebral palsy. BMC Pediatrics. 2010;10:81.

62.   Waters E DE, Boyd R, Reddihough D, Mackinnon A, Graham HK, Lo SK, Wolfe R, Stevenson R, Bjornson K, Blair E & Ravens-Sieberer U. Cerebral Palsy Quality of Life Questionnaire for Children (CP QOL-Child) manual. . Melbourne: University of Melbourne. 2013

63.   Davis E, Mackinnon A, Davern M, Boyd R, Waters E, Graham HK, et al. Description and psychometric properties of the CP QOL-Teen: a quality of life questionnaire for adolescents with cerebral palsy. Research in developmental disabilities. 2013;34(1):344-52.

64.   Waters E, Davis E, Mackinnon A, Boyd R, Graham H, Kai Lo S, et al. Psychometric properties of the quality of life questionnaire for children with CP. Developmental Medicine and Child Neurology. 2007;49(1):49-55.

65.   Solans M, Pane S, Estrada MD, Serra-Sutton V, Berra S, Herdman M, et al. Health-related quality of life measurement in children and adolescents: a systematic review of generic and disease-specific instruments. Value in health. 2008;11(4):742-64.

66.   Ratcliffe J, Couzner L, Flynn T, Sawyer M, Stevens K, Brazier J, et al. Valuing Child Health Utility 9D health states with a young adolescent sample: a feasibility study to compare best-worst scaling discrete-choice experiment, standard gamble and time trade-off methods. Appl Health Econ Health Policy. 2011;9(1):15-27.

67.   Ungar W. Economic evaluation in child health: Oxford University Press; 2009.

68.   Chen G, Ratcliffe J. A review of the development and application of generic multi-attribute utility instruments for paediatric populations. Pharmacoeconomics. 2015;33(10):1013-28.

69.   Bahrampour M, Byrnes J, Norman R, Scuffham PA, Downes M. Discrete choice experiments to generate utility values for multi-attribute utility instruments: a systematic review of methods. The European Journal of Health Economics. 2020. 21, 983–992.

70.   Young T, Yang Y, Brazier JE, Tsuchiya A, Coyne K. The first stage of developing preference-based measures: constructing a health-state classification using Rasch analysis. Quality of Life Research. 2009;18(2):253-65. doi:10.1007/s11136-008-9428-0.

71.   Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol. 2010;10:22.

72.   Von Neumann J, Morgenstern O. Theory of games and economic behavior (commemorative edition): Princeton University Press; 2007.

73.   Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. Health Services Research. 1972;7(2):118.

74.   Pullenayegum E, Xie F. Scoring the 5-level EQ-5D: can latent utilities derived from a discrete choice model be transformed to health utilities derived from time tradeoff tasks? Medical Decision Making. 2013;33(4):567-78.

75.   Bleichrodt H, Johannesson M. Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs. Journal of Health Economics. 1997;16(2):155-75.

76.   Ali S, Ronaldson S. Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. British Medical Bulletin. 2012;103(1):21-44.

77.   Stolk EA, Oppe M, Scalone L, Krabbe PF. Discrete choice modeling for the quantification of health states: the case of the EQ-5D. Value in Health. 2010;13(8):1005-13.

78.   Brazier J, Rowen D, Yang Y, Tsuchiya A. Comparison of health state utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale. European Journal of Health Economics. 2012;13(5):575-87.

79.   Mulhern B, Norman R, Street DJ, Viney R. One method, many methodological choices: a structured review of discrete-choice experiments for health state valuation. PharmacoEconomics. 2019;37(1):29-43.

80.   Thurstone LL. A law of comparative judgment. Psychological Review. 1927;34(4):273.

81.   McFadden D. Econometric models for probabilistic choice among products. Journal of Business. 1980:S13-S29.

82.   McFadden D. Conditional logit analysis of qualitative choice behavior. Frontiers in Econometrics ebPZ, editor: New York: Academic Press.; 1973.

83.   Lancaster KJ. A new approach to consumer theory. Journal of Political Economy. 1966;74(2):132-57.

84.   Mangham LJ, Hanson K, McPake B. How to do (or not to do)… Designing a discrete choice experiment for application in a low-income country. Health Policy Plan. 2009;24(2):151-8.

85.   Dolan P. Developing methods that really do value the 'Q' in the QALY. Health Economics, Policy, and Law. 2008;3(Pt 1):69-77.

86.   Slevin ML, Stubbs L, Plant HJ, Wilson P, Gregory WM, Armes PJ, et al. Attitudes to chemotherapy: comparing views of patients with cancer with those of doctors, nurses, and general public. BMJ. 1990;300(6737):1458-60.

87.   Otto RA, Dobie RA, Lawrence V, Sakai C. Impact of a laryngectomy on quality of life: perspective of the patient versus that of the health care provider. The Annals of Otology, Rhinology, and Laryngology. 1997;106(8):693-9.

88.   Ubel PA, Loewenstein G, Schwarz N, Smith D. Misimagining the unimaginable: the disability paradox and health care decision making. Health Psychology : Official Journal of the Division of Health Psychology, American Psychological Association. 2005;24(4s):S57-S62.

89.   Rowen D, Mulhern B, Banerjee S, Tait R, Watchurst C, Smith SC, et al. Comparison of general population, patient, and carer utility values for dementia health states. Medical Decision Making. 2015;35(1):68-80.

90.   Stamuli E. Health outcomes in economic evaluation: who should value health? British Medical Bulletin. 2011;97:197-210.

91.   Bray N, Noyes J, Harris N, Edwards RT. Measuring the health-related quality of life of children with impaired mobility: examining correlation and agreement between children and parent proxies. BMC Research Notes. 2017;10(1):377.

92.   Bray P, Bundy AC, Ryan MM, North KN, Everett A. Health-related quality of life in boys with Duchenne muscular dystrophy: agreement between parents and their sons. Journal of Child Neurology. 2010;25(10):1188-94.

93.   Varni JW, Burwinkle TM, Sherman SA, Hanna K, Berrin SJ, Malcarne VL, et al. Health-related quality of life of children and adolescents with cerebral palsy: hearing the voices of the children. Developmental Medicine & Child Neurology. 2005;47(9):592-7.

94.   Kluge E-HW. Resource allocation in healthcare: implications of models of medicine as a profession. Medscape General Medicine. 2007;9(1):57.

95.   Griffin S, Claxton K, Sculpher M. Decision analysis for resource allocation in health care. Journal of Health Services Research & Policy. 2008;13(3_suppl):23-30.

96.   Kularatna S, Whitty JA, Johnson NW, Jayasinghe R, Scuffham PA. A comparison of health state utility values associated with oral potentially malignant disorders and oral cancer in Sri Lanka assessed using the EQ-5D-3 L and the EORTC-8D. Health & Quality Of Life Outcomes. 2016;14:101.

97.   Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. British Medical Bulletin. 2010;96(1):5-21.

98.   Gu Y, Norman R, Viney R. Estimating health state utility values from discrete choice experiments—a QALY space model approach. Health Economics. 2014;23(9):1098-114.

99.   Krabbe PF, Devlin NJ, Stolk EA, Shah KK, Oppe M, van Hout B, et al. Multinational evidence of the applicability and robustness of discrete choice modeling for deriving EQ-5D-5L health-state values. Med Care. 2014;52(11):935-43.

100.   Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. Journal of Health Economics. 2002;21(2):271-92.

101.   Bjornson K, McLaughlin J. The measurement of health-related quality of life (HRQL) in children with cerebral palsy. European Journal of Neurology. 2001;8(s5):183-93.

102.   McCabe C, Brazier J, Gilks P, Tsuchiya A, Roberts J, O'Hagan A, et al. Using rank data to estimate health state utility models. Journal of Health Economics. 2006;25(3):418-31.

103.   Ryan M, Netten A, Skåtun D, Smith P. Using discrete choice experiments to estimate a preference-based measure of outcome—an application to social care for older people. Journal of Health Economics. 2006;25(5):927-44.

104.   Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. Medical Decision Making. 2015;35(3):328-40.

105.   Lancsar E, Wildman J, Donaldson C, Ryan M, Baker R. Deriving distributional weights for QALYs through discrete choice experiments. Journal of Health Economics. 2011;30(2):466-78.

106.   Stolk EA, Oppe M, Scalone L, Krabbe PF. Discrete choice modeling for the quantification of health states: the case of the EQ-5D. Value in Health. 2010;13(8):1005-13.

107.   Brazier J, Rowen D, Yang Y, Tsuchiya A. Comparison of health state utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale. European Journal of Health Economics. 2012;13(5):575-87.

108.   Ludwig K, von der Schulenburg JG, Greiner W. Valuation of the EQ-5D-5L with composite time trade-off for the German population - an exploratory study. Health & Quality of Life Outcomes. 2017;15(1):39. doi:https://dx.doi.org/10.1186/s12955-017-0617-9.

109.   Craig BM, Busschbach JJV. Towards a more universal approach in health valuation. Health economics. 2011;20(7):10.1002/hec.650.

110.   Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics. 2014;32(9):883-902.

111.   Mulhern B, Norman R, Lorgelly P, Lancsar E, Ratcliffe J, Brazier J, et al. Is dimension order important when valuing health states using discrete choice experiments including duration? Pharmacoeconomics [Internet]. 2016:[1-13 pp.]. Available from: https://link.springer.com/article/10.1007%2Fs40273-016-0475-z.

112.   King MT, Viney R, Simon Pickard A, Rowen D, Aaronson NK, Brazier JE, et al. Australian utility weights for the EORTC QLU-C10D, a multi-attribute utility instrument derived from the Cancer-Specific Quality of Life Questionnaire, EORTC QLQ-C30. PharmacoEconomics. 2018;36(2):225-38.

113.   Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. Journal of Health Economics. 2012;31(1):306-18.

114.   Burr JM, Kilonzo M, Vale L, Ryan M. Developing a preference-based Glaucoma Utility Index using a discrete choice experiment. Optom Vis Sci. 2007;84(8):797-808.

115.   Ratcliffe J, Brazier J, Tsuchiya A, Symonds T, Brown M. Using DCE and ranking data to estimate cardinal values for health states for deriving a preference-based single index from the sexual quality of life questionnaire. Health Economics. 2009;18(11):1261-76.

116.   Moher D, Liberati A, Tetzlaff J, Altman D. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of Internal Medicine . 2009;151(4):264-9.

117.   Xie F, Pickard AS, Krabbe PF, Revicki D, Viney R, Devlin N, et al. A Checklist for Reporting Valuation Studies of Multi-Attribute Utility-Based Instruments (CREATE). Pharmacoeconomics. 2015;33(8):867-77.

118.   Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value in Health. 2011;14(4):403-13.

119.   de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Economics. 2012;21(2):145-72.

120.   Selivanova A, Buskens E, Krabbe PFM. Head-to-head comparison of EQ-5D-3L and EQ-5D-5L health values. PharmacoEconomics. 2018;36(6):715-25.

121.   Norman R, Viney R, Aaronson N, Brazier J, Cella D, Costa D, et al. Using a discrete choice experiment to value the QLU-C10D: feasibility and sensitivity to presentation format. Quality of Life Research. 2016;25(3):637-49.

122.   Norman R, Kemmler G, Viney R, Pickard AS, Gamper E, Holzner B, et al. Order of presentation of dimensions does not systematically bias utility weights from a discrete choice experiment. Value in Health. 2016;19(8):1033-8.

123.   Gamper E-M, Holzner B, King MT, Norman R, Viney R, Nerich V, et al. Test-retest reliability of discrete choice experiment for valuations of QLU-C10D health states. Value in Health. 2018;21(8):958-66.

124.   Huynh E, Coast J, Rose J, Kinghorn P, Flynn T. Values for the ICECAP-Supportive Care Measure (ICECAP-SCM) for use in economic evaluation at end of life. Social Science & Medicine. 2017;189:114-28.

125.   Xie F, Pullenayegum E, Pickard AS, Ramos Goñi JM, Jo Mw, Igarashi A, et al. Transforming latent utilities to health utilities: east does not meet west. Health Economics. 2017;26(12):1524-33.

126.   Krabbe PFM, Stolk EA, Devlin NJ, Xie F, Quik EH, Pickard AS. Head-to-head comparison of health-state values derived by a probabilistic choice model and scores on a visual analogue scale. European Journal of Health Economics. 2017;18(8):967-77.

127.   Scalone L, Stalmeier PF, Milani S, Krabbe PF. Values for health states with different life durations. European Journal of Health Economics. 2015;16(9):917-25.

128.   Ratcliffe J, Flynn T, Terlich F, Stevens K, Brazier J, Sawyer M. Developing adolescent-specific health state values for economic evaluation: an application of profile case best-worst scaling to the Child Health Utility 9D. Pharmacoeconomics. 2012;30(8):713-27.

129.   Prosser LA, Payne K, Rusinak D, Shi P, Messonnier M. Using a discrete choice experiment to elicit time trade-off and willingness-to-pay amounts for influenza health-related quality of life at different ages. Pharmacoeconomics. 2013;31(4):305-15.

130.   Purba F, Hunfeld J, Iskandarsyah A, Fitriana T, Sadarjoen S, Ramos-Goñi J, et al. The Indonesian EQ-5D-5L Value Set. PharmacoEconomics. 2017;35(11):1153-65.

131.   Eckert J, Lichters M, Piedmont S, Vogt B, Robra BP. Case vignettes based on EQ-5D to elicit stated preferences for health services utilization from the insurees' perspective. BMC Health Serv Res. 2015;15:481.

132.   Jonker MF, Donkers B, de Bekker-Grob EW, Stolk EA. Advocating a paradigm shift in health-state valuations: the estimation of time-preference corrected QALY tariffs. Value in Health. 2018;21(8):993-1001.

133.   Craig BM, Rand K. Choice defines QALYs: a US valuation of the EQ-5D-5L. Medical Care. 2018;56(6):529-36.

134.   Dolan P, Kind P. Inconsistency and health state valuations. Social Science & Medicine. 1996;42(4):609-15.

135.   Ramos-Goni JM, Rivero-Arias O, Errea M, Stolk EA, Herdman M, Cabases JM. Dealing with the health state 'dead' when using discrete choice experiments to obtain values for EQ-5D-5L health states. European Journal of Health Economics. 2013;14 Suppl 1:S33-S42.

136.   Ramos-Goñi JM, Pinto-Prades JL, Oppe M, Cabasés JM, Serrano-Aguilar P, Rivero-Arias O. Valuation and modeling of EQ-5D-5L health states using a hybrid approach. Medical Care. 2017;55(7):e51-e8.

137.   Bansback N, Hole AR, Mulhern B, Tsuchiya A. Testing a discrete choice experiment including duration to value health states for large descriptive systems: addressing design and sampling issues. Social Science & Medicine. 2014;114:38-48.

138.	Mulhern B, Shah K, Janssen MF, Longworth L, Ibbotson R. Valuing health using time trade-off and discrete choice experiment methods: does dimension order impact on health state values? Value in Health. 2016;19(2):210-217.

139.	Shiroiwa T, Ikeda S, Noto S, Igarashi A, Fukuda T, Saito S, et al. Comparison of value set based on DCE and/or TTO data: scoring for EQ-5D-5L health states in Japan. Value in Health. 2016;19(5):648-54.

140.	Versteegh MM, Brouwer WBF. Patient and general public preferences for health states: A call to reconsider current guidelines. Social Science & Medicine. 2016;165:66-74.

141.	Huber J, Zwerina K. The importance of utility balance in efficient choice designs. Journal of Marketing Research. 1996:307-17.

142.	Vanniyasingam T, Cunningham CE, Foster G, Thabane L. Simulation study to determine the impact of different design features on design efficiency in discrete choice experiments. BMJ Open. 2016;6(7). e011985. doi: 10.1136/bmjopen-2016-011985

143.	Hensher DA, Rose JM, Greene WH. Applied Choice Analysis.3rd ed; Cambridge University Press; 2015.

144.	Kitamura R, Yoshii T, Yamamoto T. Expanding sphere of travel behaviour research: selected papers from the 11th International Conference On Travel Behaviour Research: Emerald Group Publishing; 2009.

145.	De Martino B, Harrison NA, Knafo S, Bird G, Dolan RJ. Explaining enhanced logical consistency during decision making in autism. Journal of Neuroscience. 2008;28(42):10746-50.

146.	Arrow KJ. Risk perception in psychology and economics. Economic Inquiry. 1982;20(1):1-9.

147.	Norman R, Cronin P, Viney R. A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states. Appl Health Econ Health Policy. 2013;11(3):287-98.

148.	Mulhern B, Bansback N, Hole AR, Tsuchiya A. Using discrete choice experiments with duration to model EQ-5D-5L health state preferences. Medical Decision Making. 2017;37(3):285-97.

149.	Lancsar E, Louviere J. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Economics. 2006;15(8):797-811.

150.	Hakim Z. Modelling the EuroQol data: a comparison of discrete choice conjoint and conditional preference modelling. Health Economics. 1999;8(2):103-16.

151.	Lamers LM. The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value sets. Medical Care. 2007;45(3):238-44.

152.	Robinson A, Spencer A, Moffatt P. A framework for estimating health state utility values within a discrete choice experiment: modeling risky choices. Medical Decision Making. 2015;35(3):341-50.

153.	Attema AE, Edelaar-Peeters Y, Versteegh MM, Stolk EA. Time trade-off: one methodology, different methods. The European Journal of Health Economics. 2013;14(Suppl 1):S53-S64.

154.	Hauber AB, González JM, Groothuis-Oudshoorn CGM, Prior T, Marshall DA, Cunningham C, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR Conjoint Analysis Good Research Practices Task Force. Value in Health. 2016;19(4):300-15.

155.	Rosenbaum P, Paneth N, Leviton A, Goldstein M, Bax M, Damiano D, et al. A report: the definition and classification of cerebral palsy April 2006. Developmental Medicine & Child Neurology. 2007;109:8.

156.	Oskoui M, Coutinho F, Dykeman J, Jetté N, Pringsheim T. An update on the prevalence of cerebral palsy: a systematic review and meta-analysis. Developmental Medicine & Child Neurology. 2013;55(6):509-19.

157.	Graham HK. Absence of reference to progressive musculoskeletal pathology in definition of cerebral palsy. Developmental Medicine and Child Neurology. 2006;48(1):78-9.

158.	Zilberberg MD, Shorr AF. Understanding cost-effectiveness. Clinical Microbiology and Infection. 2010;16(12):1707-12.

159. Novak I, McIntyre S, Morgan C, Campbell L, Dark L, Morton N, et al. A systematic review of interventions for children with cerebral palsy: state of the evidence. Developmental Medicine & Child Neurology. 2013;55(10):885-910.

160. Torrance GW. Utility approach to measuring health-related quality of life. Journal of Chronic Diseases. 1987;40(6):593-600.

161. Chen KL, Wang HY, Tseng MH, Shieh JY, Lu L, Yao KP, et al. The Cerebral Palsy Quality of Life for Children (CP QOL-Child): evidence of construct validity. Research in developmental disabilities. 2013;34(3):994-1000.

162. Atasavun Uysal S, Duger T, Elbasan B, Karabulut E, Toylan I. Reliability and validity of the Cerebral Palsy Quality of Life Questionnaire in the Turkish population. Perceptual and motor skills. 2016;122(1):150-64.

163. Lim HK, Ko J. Reliability and Validity of the Cerebral Palsy Quality of Life Questionnaire in the South Korean Population. Journal of the Ergonomics Society of Korea [Internet]. 2018; 37 (6). doi:http://dx.doi.org/10.5143/JESK.2018.37.6.733.

164. El-Weshahi H, Mohamed M, Abd-Elghany H, Omar T, Azzawi A. Psychometric properties of a translated arabic version of cerebral palsy-quality of life questionnaire: primary caregiver form. Alexandria Journal of Pediatrics. 2017;30(2):53-60.

165. Wang HY, Cheng CC, Hung JW, Ju YH, Lin JH, Lo SK. Validating the Cerebral Palsy Quality of Life for Children (CP QOL-Child) questionnaire for use in Chinese populations. Neuropsychological rehabilitation. 2010;20(6):883-98.

166. Das S, Aggarwal A, Roy S, Kumar P. quality of life in Indian children with Cerebral palsy using Cerebral Palsy Quality of Life questionnaire. Journal of Pediatric Neurosciences. 2017;12(3):251-4.

167. Davis E, Shelly A, Waters E, Davern M. Measuring the quality of life of children with cerebral palsy: comparing the conceptual differences and psychometric properties of three instruments. Developmental Medicine & Child Neurology. 2010;52(2):174-80.

168. Comans T, Mihala G, Sakzewski L, Boyd RN, Scuffham P. The cost-effectiveness of a web-based multimodal therapy for unilateral cerebral palsy: the Mitii randomized controlled trial. Developmental Medicine & Child Neurology. 2017;59(7):756-61.

169. Hoving MA, Evers S, Ament A, Van Raak E, Vles J, Spasticity. DSGoC. Intrathecal baclofen therapy in children with intractable spastic cerebral palsy: a cost-effectiveness analysis. Developmental Medicine & Child Neurology. 2008;50(6):450-5.

170. Christensen R, Macintosh A, Switzer L, Fehlings D. Change in pain status in children with cerebral palsy. Developmental Medicine & Child Neurology. 2017;59(4):374-9.

171. Kennes J, Rosenbaum P, Hanna SE, Walter S, Russell D, Raina P, et al. Health status of school-aged children with cerebral palsy: information from a population-based sample. Developmental Medicine & Child Neurology. 2002;44(4):240-7.

172. Penner M, Xie WY, Binepal N, Switzer L, Fehlings D. Characteristics of pain in children and youth with cerebral palsy. Pediatrics. 2013;132(2):e407-e13.

173. Rosenbaum PL, Livingston MH, Palisano RJ, Galuppi BE, Russell DJ. Quality of life and health-related quality of life of adolescents with cerebral palsy. Developmental Medicine & Child Neurology. 2007;49(7):516-21.

174. Young NL, Rochon TG, McCormick A, Law M, Wedge JH, Fehlings D. The health and quality of life outcomes among youth and young adults with cerebral palsy. Archives of Physical Medicine and Rehabilitation. 2010;91(1):143-8.

175. Mpundu-Kaambwa C, Chen G, Huynh E, Russo R, Ratcliffe J. A review of preference-based measures for the assessment of quality of life in children and adolescents with cerebral palsy. Quality of Life Research. 2018: 27(7):1781-1799. doi: 10.1007/s11136-018-1837-0.

176. Young TA, Yang Y, Brazier JE, Tsuchiya A. The use of rasch analysis in reducing a large condition-specific instrument for preference valuation: the case of moving from AQLQ to AQL-5D. Medical Decision Making. 2011;31(1):195-210.

177.   King MT, Costa DS, Aaronson NK, Brazier JE, Cella DF, Fayers PM, et al. QLU-C10D: a health state classification system for a multi-attribute utility measure based on the EORTC QLQ-C30. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2016;25(3):625-36.

178.   Goodwin E, Green C, Spencer A. Estimating a preference-based index for an eight-dimensional health state classification system for Multiple Sclerosis. Value in Health. 2015;18(8):1025-36.

179.   Waters E, Davis E, Boyd R, Reddihough D, Mackinnon A, Graham HK, Lo SK, Wolfe R, Stevenson R, Bjornson K, Blair E & Ravens-Sieberer U. Cerebral Palsy Quality of Life Questionnaire for Children (CP QOL-Child) Manual. Melbourne, Australia: University of Melbourne,; 2013

180.   Boyd RN, Baque E, Piovesana A, Ross S, Ziviani J, Sakzewski L, et al. Mitii™ ABI: study protocol of a randomised controlled trial of a web-based multi-modal training program for children and adolescents with an Acquired Brain Injury (ABI). BMC Neurology. 2015;15(1):140.

181.   Boyd RN, Sakzewski L, Ziviani J, Abbott DF, Badawy R, Gilmore R, et al. INCITE: A randomised trial comparing constraint induced movement therapy and bimanual training in children with congenital hemiplegia. BMC Neurology. 2010;10(1):4.

182.   Boyd RN, Jordan R, Pareezer L, Moodie A, Finn C, Luther B, et al. Australian Cerebral Palsy Child Study: protocol of a prospective population based study of motor and brain development of preschool aged children with cerebral palsy. BMC Neurology. 2013;13(1):57.

183.   Brazier JE, Rowen D, Mavranezouli I, Tsuchiya A, Young T, Yang Y, et al. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). Health Technology Assessment. 2012;16(32):1-114.

184.   Kaiser HF. A second generation little jiffy. Psychometrika. 1970;35(4):401-15.

185.   Bartlett MS. Tests of significance in factor analysis. British Journal of Mathematical and Statistical Psychology. 1950;3(2):77-85.

186.   Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL. Multivariate data analysis: Prentice Hall Upper Saddle River, NJ; 1998.

187.   Tabachnick BG, Fidell LS. Using multivariate statistics: Allyn & Bacon/Pearson Education; 2007.

188.   Jolliffe IT. Principal component analysis and factor analysis. Principal component analysis: Springer; 1986. p. 115-28.

189.   Kaiser HF. The application of electronic computers to factor analysis. Educational and Psychological Measurement. 1960;20(1):141-51.

190.   Williams B, Onsman A, Brown T. Exploratory factor analysis: a five-step guide for novices. Australasian Journal of Paramedicine. 2010;8(3).

191.   Young TA, Yang Y, Brazier JE, Tsuchiya A. The use of Rasch analysis in reducing a large condition-specific instrument for preference valuation: the case of moving from AQLQ to AQL-5D. Medical Decision Making. 2011;31(1):195-210.

192.   Jabrayilov R, Emons WH, Sijtsma K. Comparison of classical test theory and item response theory in individual change assessment. Applied Psychological Measurement. 2016;40(8):559-72.

193.   Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis and Rheumatism. 2007;57(8):1358-62.

194.   Miles MB, Huberman AM. Qualitative data analysis: A sourcebook of new methods. 2nd ed; Sage publications; 1984.

195.   Carnahan KD, Arner M, Hägglund G. Association between gross motor function (GMFCS) and manual ability (MACS) in children with cerebral palsy. A population-based study of 359 children. BMC musculoskeletal disorders. 2007;8:50-7. doi:10.1186/1471-2474-8-50.

196.   Newman CJ, O'Regan M, Hensey O. Sleep disorders in children with cerebral palsy. Developmental Medicine & Child Neurology. 2006;48(7):564-8.

197.   Schwartz L, Engel JM, Jensen MP. Pain in persons with cerebral palsy. Archives of Physical Medicine and Rehabilitation. 1999;80(10):1243-6.
198.   Bjornson KF, Belza B, Kartin D, Logsdon R, McLaughlin J, Thompson EA. The relationship of physical activity to health status and quality of life in cerebral palsy. Pediatric Physical Therapy : the Official Publication of the Section on Pediatrics of the American Physical Therapy Association. 2008;20(3):247-53.
199.   Romeo DM, Brogna C, Quintiliani M, Baranello G, Pagliano E, Casalino T, et al. Sleep disorders in children with cerebral palsy: neurodevelopmental and behavioral correlates. Sleep Medicine. 2014;15(2):213-8.
200.   Bautista M, Whittingham K, Edwards P, Boyd RN. Psychometric properties of parent and child reported sleep assessment tools in children with cerebral palsy: a systematic review. Developmental Medicine & Child Neurology. 2018;60(2):162-72.
201.   Varni JW, Burwinkle, T.M., Berrin, S.J., Sherman, S.A., Artavia, K., Malcarne, V.L., & Chambers, H.G. The PedsQL™ in pediatric cerebral palsy: Reliability, validity, and sensitivity of the Generic Core Scales and Cerebral Palsy Module.Developmental Medicine & Child Neurology . 2006;48:442-9. doi: 10.1017/S001216220600096X
202.   Narayanan UG, Fehlings DL, Campbell K, Weir S, Kiran S. Caregiver Priorities and Child Health Index of Life with Disabilities: development and initial validation of a disease-specific measure of health status and well being in children with severe disabilities due to cerebral palsy. Developmental Medicine & Child Neurology. 2004;46:6.
203.   Simeoni MC, Schmidt S, Muehlan H, Debensason D, Bullinger M. Field testing of a European quality of life instrument for children and adolescents with chronic conditions: the 37-item DISABKIDS Chronic Generic Module. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2007;16(5):881-93.
204.   McCoy RN, Blasco PA, Russman BS, O'Malley JP. Validation of a care and comfort hypertonicity questionnaire. Developmental Medicine & Child Neurology. 2006;48(3):181-7.
205.   Eliasson AC, Krumlinde-Sundholm L, Rosblad B, Beckung E, Arner M, Ohrvall AM, et al. The Manual Ability Classification System (MACS) for children with cerebral palsy: scale development and evidence of validity and reliability. Developmental Medicine & Child Neurology. 2006;48(7):549-54.
206.   Goodwin E, Green C. A quality-adjusted life-year measure for Multiple Sclerosis: developing a patient-reported health state classification system for a Multiple Sclerosis-specific preference-based measure. Value in Health. 2015;18(8):1016-24.
207.   Davis E, Shelly A, Waters E, Mackinnon A, Reddihough D, Boyd R, et al. Quality of life for adolescents with cerebral palsy: Perspectives of adolescents and parents. Developmental Medicine & Child Neurology. 2009;51:193.
208.   Shelly A, Davis E, Waters E, Mackinnon A, Reddihough D, Boyd R, et al. The relationship between quality of life and functioning for children with cerebral palsy. Developmental Medicine & Child Neurology. 2008;50(3):199-203.
209.   Neumann PJ, Willke RJ, Garrison LP, Jr. A health economics approach to US value assessment frameworks-introduction: An ISPOR Special Task Force Report [1]. Value in Health. 2018;21(2):119-23.
210.   Birch S, Gafni A. Cost effectiveness/utility analyses: Do current decision rules lead us to where we want to be? Journal of Health Economics. 1992;11(3):279-96.
211.   Jakubiak-Lasocka J, Jakubczyk M. Cost-effectiveness versus cost-utility analyses: what are the motives behind using each and how do their results differ?—A Polish example. Value in Health Regional Issues. 2014;4:66-74.
212.   Utility[online]. York: York Health Economics Consortium; (2016) [Available from: http://www.yhec.co.uk/glossary/utility/.
213.   Ryan M, Scott DA, Reeves C, Bate A, Russell E, Napper M, et al. Eliciting public preferences for healthcare: a systematic review of techniques. Health Technology Assessment2001. p. 1-186.

214.	Mulhern B, Rowen D, Brazier J, Smith S, Romeo R, Tait R, et al. Development of DEMQOL-U and DEMQOL-PROXY-U: generation of preference-based indices from DEMQOL and DEMQOL-PROXY for use in economic evaluation. Health Technology Assessment. 2013;17(5).
215.	Bahrampour M, Downes M, Boyd RN, Scuffham PA, Byrnes J. Using Rasch and factor analysis to develop a proxy-reported health state classification (descriptive) system for Cerebral Palsy. Disability and Rehabilitation. 2020:1-9.
216.	Waters E DE, Boyd R, Reddihough D, Mackinnon A, Graham HK, Lo SK, Wolfe R, Stevenson R, Bjornson K,, U. BER-S. Cerebral Palsy Quality of Life Questionnaire for Children (CP QOL-Child) Manual. Melbourne: University of Melbourne.; 2013
217.	Peeters Y, Vliet Vlieland TPM, Stiggelbout AM. Focusing illusion, adaptation and EQ-5D health state descriptions: the difference between patients and public. Health Expect. 2012;15(4):367-78.
218.	Stamuli E. Health outcomes in economic evaluation: who should value health? British Medical Bulletin. 2011;97(1):197-210.
219.	L. KR, von WD. A prescriptive risk framework for individual health and safety decisions. Risk Analysis. 1991;11(3):523-33.
220.	Ryan M. Discrete choice experiments in health care: NICE should consider using them for patient centred evaluations of technologies. BMJ. 2004;328(7436):360.
221.	Mangham LJ, Hanson K, McPake B. How to do (or not to do)… Designing a discrete choice experiment for application in a low-income country. Health Policy and Planning. 2009;24(2):151-8.
222.	Ryan M, Netten A, Skåtun D, Smith P. Using discrete choice experiments to estimate a preference-based measure of outcome—An application to social care for older people. Journal of Health Economics. 2006;25(5):927-44.
223.	Scalone L, Stalmeier PF, Milani S, Krabbe PF. Values for health states with different life durations. European Journal of Health Economics. 2015;16(9):917-25.
224.	ChoiceMetrics. User manual & reference guide.Ngene. Sydney, Australia: ChoiceMetrics; 2014.
225.	Reed Johnson F, Lancsar E, Marshall D, Kilambi V, Muhlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value Health. 2013;16(1):3-13.
226.	Hawthorne G RJ, & Osbourne R. . The Assessment of Quality of life (AQoL) instrument: a psychometric measure of health-related quality of life. Quality of life research. 1999;8(3):209-24.
227.	Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. Developmental Medicine & Child Neurology. 1997;39(4):214-23.
228.	Morris C, Kurinczuk JJ, Fitzpatrick R, Rosenbaum PL. Reliability of the manual ability classification system for children with cerebral palsy. Developmental Medicine & Child Neurology. 2006;48(12):950-3.
229.	Eliasson A-C, Krumlinde-Sundholm L, Rösblad B, Beckung E, Arner M, Öhrvall A-M, et al. The Manual Ability Classification System (MACS) for children with cerebral palsy: scale development and evidence of validity and reliability. Developmental Medicine & Child Neurology. 2006;48(7):549-54.
230.	Piscitelli D, Vercelli S, Meroni R, Zagnoni G, Pellicciari L. Reliability of the gross motor function classification system and the manual ability classification system in children with cerebral palsy in Tanzania. Developmental Neurorehabilitation. 2017:1-7.
231.	SurveyEngine GmbH. SurveyEngine decision modelling technology 2017. [Available from: http://trial.surveyengine.com/about.html.
232.	Wittenberg E, Prosser LA. Ordering errors, objections and invariance in utility survey responses: a framework for understanding who, why and what to do. Appl Health Econ Health Policy. 2011;9(4):225-41.
233.	Reed Johnson F, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: Report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value in Health. 2013;16(1):3-13.

234.   Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making. Pharmacoeconomics. 2008;26(8):661-77.
235.   Thorrington D, Eames K. Measuring health utilities in children and adolescents: a systematic review of the literature. PloS One. 2015;10(8):e0135672. https://doi.org/10.1371/journal.pone.0135672
236.   Varni JW, Limbers CA, Burwinkle TM. Parent proxy-report of their children's health-related quality of life: an analysis of 13,878 parents' reliability and validity across age subgroups using the PedsQL™ 4.0 Generic Core Scales. Health & Quality Of Life Outcomes. 2007;5(1):2. doi: 10.1186/1477-7525-5-2
237.   Hauber AB, González JM, Groothuis-Oudshoorn CG, Prior T, Marshall DA, Cunningham C, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR Conjoint Analysis Good Research Practices Task Force. Value in Health. 2016;19(4):300-15.
238.   Hess S, Train K. Correlation and scale in mixed logit models. Journal of Choice Modelling. 2017;23:1-8.
239.   Flynn TN, Louviere JJ, Peters TJ, Coast J. Best–worst scaling: what it can do for health care research and how to do it. Journal of Health Economics. 2007;26(1):171-89.
240.   Nord E. Health state values from multiattribute utility instruments need correction. Annals of Medicine. 2001;33(5):371-4.
241.   Torrance GW, Keresteci MA, Casey RW, Rosner AJ, Ryan N, Breton MC. Development and initial validation of a new preference-based disease-specific health-related quality of life instrument for erectile function. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2004;13(2):349-59.
242.   Dowie J. Decision validity should determine whether a generic or condition-specific HRQOL measure is used in health care decisions. Health Econ. 2002;11(1):1-8.
243.   Stolk EA, Busschbach JJ. Validity and feasibility of the use of condition-specific outcome measures in economic evaluation. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2003;12(4):363-71.
244.   Mangham LJ, Hanson K, McPake B. How to do (or not to do) … Designing a discrete choice experiment for application in a low-income country. Health Policy and Planning. 2008;24(2):151-8.
245.   Viney R, Norman R, Brazier J, Cronin P, King MT, Ratcliffe J, et al. An Australian discrete choice experiment to value EQ-5D health states. Health Economics. 2014;23(6):729-42.
246.   Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. Journal of Health Economics. 2012;31(1):306-18.
247.   Weinstein MC, Torrance G, McGuire A. QALYs: the basics. Value in Health. 2009;12:S5-S9.
248.   Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. Health Economics. 1996;5(2):141-54.
249.   Norman R, Cronin P, Viney R. A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states. Applied Health Economics and Health Policy. 2013;11(3):287-98.
250.   Kuha J. AIC and BIC: Comparisons of assumptions and performance. Sociological Methods & Research. 2004;33(2):188-229.
251.   Bleichrodt H, Wakker P, Johannesson M. Characterizing QALYs by risk neutrality. Journal of Risk and Uncertainty. 1997;15(2):107-14.
252.   Hidecker MJC, Paneth N, Rosenbaum PL, Kent RD, Lillie J, Eulenberg JB, et al. Developing and validating the Communication Function Classification System for individuals with cerebral palsy. Developmental Medicine & Child Neurology. 2011;53(8):704-10.
253.   Shah KK, Mulhern B, Longworth L, Janssen MF. Views of the UK general public on important aspects of health not captured by EQ-5D. Patient - Patient-Centered Outcomes Research. 2017;10(6):701-9.
254.   Furlong WJ, Feeny DH, Torrance GW, Barr RD. The Health Utilities Index (HUI®) system for assessing health-related quality of life in clinical studies. Annals of Medicine. 2001;33(5):375-84.

255.   Medvedev ON, Landhuis CE. Exploring constructs of well-being, happiness and quality of life. PeerJ. 2018;6:e4903-e. doi: 10.7717/peerj.4903

256.   Blanchflower DG, Oswald AJ. Well-being over time in Britain and the USA. Journal of Public Economics. 2004;88(7-8):1359-86.

257.   Schünemann J, Strulik H, Trimborn T. Going from bad to worse: adaptation to poor health health spending, longevity, and the value of life. Journal of Economic Behavior & Organization. 2017;140:130-46.

258.   Norman R, Cronin P, Viney R, King M, Street D, Ratcliffe J. International comparisons in valuing EQ-5D health states: a review and analysis. Value in Health. 2009;12(8):1194-200.

259.   Fedrizzi E, Pagliano E, Andreucci E, Oleari G. Hand function in children with hemiplegic cerebral palsy: prospective follow-up and functional outcome in adolescence. Developmental Medicine & Child Neurology. 2003;45(2):85-91.

260.   Ferreira LN, Ferreira PL, Pereira LN, Rowen D, Brazier JE. Exploring the consistency of the SF-6D. Value in Health. 2013;16(6):1023-31

261.   Newman CJ, O'Regan M, Hensey O. Sleep disorders in children with cerebral palsy. Developmental Medicine & Child Neurology. 2006;48(7):564-8.

262.   Lim S, Jonker MF, Oppe M, Donkers B, Stolk E. Severity-Stratified Discrete choice experiment designs for health state evaluations. PharmacoEconomics. 2018;36(11):1377-89.

263.   Helgesson G, Ernstsson O, Åström M, Burström K. Whom should we ask? A systematic literature review of the arguments regarding the most accurate source of information for valuation of health states. Quality of Life Research. 2020:1-18.

264.   Rowen D, Azzabi Zouraq I, Chevrou-Severac H, van Hout B. International regulations and recommendations for utility data for health technology assessment. PharmacoEconomics. 2017;35(1):11-19.

265.   Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: past, present and future. Applied Health Economics and Health Policy. 2017;15(2):127-37.

266.   Herdman M, Kerr C, Pavesi M, Garside J, Lloyd A, Cubi-Molla P, et al. Testing the validity and responsiveness of a new cancer-specific health utility measure (FACT-8D) in relapsed/refractory mantle cell lymphoma, and comparison to EQ-5D-5L. Journal of Patient-Reported Outcomes. 2020;4(1):22-34.

267.   Richardson JH, G The Australian Quality of Life (AQOL) instrument: psychometric properties of the descriptive system and initial validation. 19th Australian Conference of Health Economists; 1/07/97 - 1/07/97 Melbourne, Vic, Australia1998. 343-6.

268.   Hawthorne G, Richardson J, Osborne R. The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of health-related quality of life. Quality of Life Research. 1999;8(3):209-24.

269.   Field A. Discovering statistics using IBM SPSS statistics: Sage; 2013.

270.   Ghauri P, Grønhaug K, Strange R. Research methods in business studies: Cambridge University Press; 2020.

271.   Cohen J. A power primer. Psychological Bulletin. 1992;112(1):155-9.

272.   Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Economics. 2004;13(9):873-84.

273.   Lamers LM, Bouwmans CAM, van Straten A, Donker MCH, Hakkaart L. Comparison of EQ-5D and SF-6D utilities in mental health patients. Health Economics. 2006;15(11):1229-36.

274.   Richardson J, Khan MA, Iezzi A, Maxwell A. Comparing and explaining differences in the magnitude, content, and sensitivity of utilities predicted by the EQ-5D, SF-6D, HUI 3, 15D, QWB, and AQoL-8D multiattribute utility instruments. Medical Decision Making. 2015;35(3):276-91.

275.   Rowen D, Young T, Brazier J, Gaugris S. Comparison of generic, condition-specific, and mapped health state utility values for Multiple Myeloma Cancer. Value in Health. 2012;15(8):1059-68.

276.   Hawthorne G, Richardson J, Day N, Osborne R, McNeil H. Construction and utility scaling of the Assessment of Quality of Life (AQoL) instrument. Melbourne: Centre for Health Program Evaluation; 2000.

277.	Bahrampour M, Norman R, Byrnes J, Downes M, Scuffham PA. Utility Values for the CP-6D, a Cerebral Palsy-Specific Multi-Attribute Utility Instrument, Using a Discrete Choice Experiment. The Patient - Patient-Centered Outcomes Research. 2020. https://doi.org/10.1007/s40271-020-00468-x

278.	Cohen J. Statistical power analysis for the behavioral sciences.Hillsdale (NJ): Lawrence Erlbaum Associates. 1988.

279.	Adams MA, Conway TL. Eta Squared. In: Michalos AC, editor. Encyclopedia of Quality of Life and Well-Being Research. Dordrecht: Springer Netherlands; 2014. p. 1965-6.

280.	Bansback N, Sun H, Guh DP, Li X, Nosyk B, Griffin S, et al. Impact of the recall period on measuring health utilities for acute events. Health Economics. 2008;17(12):1413-9.

281.	Mulhern B, Pink J, Rowen D, Borghs S, Butt T, Hughes D, et al. Comparing generic and condition-specific preference-based measures in Epilepsy: EQ-5D-3L and NEWQOL-6D. Value in Health. 2017;20(4):687-93.

282.	Kularatna S, Byrnes J, Chan YK, Carrington MJ, Stewart S, Scuffham PA. Comparison of contemporaneous responses for EQ-5D-3L and Minnesota Living with Heart Failure; a case for disease specific multiattribute utility instrument in cardiovascular conditions. International Journal of Cardiology. 2017;227:172-6.

283.	Löwe B, Wahl I, Rose M, Spitzer C, Glaesmer H, Wingenfeld K, et al. A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. Journal of Affective Disorders. 2010;122(1):86-95.

284.	Richardson J, Iezzi A, Khan MA, Chen G, Maxwell A. Measuring the sensitivity and construct validity of 6 utility instruments in 7 disease areas. Medical Decision Making. 2015;36(2):147-59.

285.	Hawthorne G, Korn S, Richardson J. Population norms for the AQoL derived from the 2007 Australian National Survey of Mental Health and Wellbeing. Australian and New Zealand Journal of Public Health. 2013;37(1):7-16.

286.	Vargus-Adams J. Health-related quality of life in childhood cerebral palsy. Archives of physical medicine and rehabilitation. 2005;86(5):940-5.

287.	Okurowska–Zawada B, Kułak W, Wojtkowski J, Sienkiewicz D, Paszko-Patej G. Quality of life of parents of children with cerebral palsy. Progress in Health Sciences. 2011;1(1):116-23.

288.	Eker L, Tüzün EH. An evaluation of quality of life of mothers of children with cerebral palsy. Disabil Rehabil. 2004;26(23):1354-9.

289.	van Velden ME, Severens JL, Novak A. Economic evaluations of healthcare programmes and decision making: the influence of economic evaluations on different healthcare decision-making levels. Pharmacoeconomics. 2005;23(11):1075-82.

290.	National Institute for Health and Care Excellence (NICE). Guide to the methods of technology appraisal 2013. Available from: https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781.

291.	Guidelines for the economic evaluation of health technologies: Canada (Ottawa): Canadian Agency for Drugs and Technologies in Health; 2006.

292.	Mathes T, Jacobs E, Morfeld JC, Pieper D. Methods of international health technology assessment agencies for economic evaluations-a comparative analysis. BMC Health Serv Res. 2013(13):371.

293.	Medical Services Advisory Committee. Technical guidelines for preparing assessment reports for the medical services advisory committee–medical service type: Therapeutic (version 2.0). Department of Health, Australian Government; 2016.

294.	Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the economic evaluation of health care programmes. 3rd ed. Oxford: Oxford University Press; 2005.

295.	Stevens K, Palfreyman S. The use of qualitative methods in developing the descriptive systems of preference-based measures of health-related quality of life for use in economic evaluation. Value in Health. 2012;15(8):991-8.

296.   Rowen D, Brazier J, Ara R, Zouraq IA. The role of condition-specific preference-based measures in health technology assessment. Pharmacoeconomics. 2017;35(1):33-41.
297.   Gold M, Franks P, Erickson P. Assessing the health of the nation. The predictive validity of a preference-based measure and self-rated health. Medical Care. 1996;34(2):163-77.
298.   Petrou S, Kupek E. Estimating preference-based health utilities index mark 3 utility scores for childhood conditions in England and Scotland. Medical Decision Making. 2009;29(3):291-303.
299.   Carroll AE, Downs SM. Improving decision analyses: parent preferences (utility values) for pediatric health outcomes. The Journal of Pediatrics. 2009;155(1):21-5.
300.   Whitehurst DGT, Mittmann N, Noonan VK, Dvorak MF, Bryan S. Health state descriptions, valuations and individuals' capacity to walk: a comparative evaluation of preference-based instruments in the context of spinal cord injury. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2016;25(10):2481-96.
301.   Nguyen K-H, Mulhern B, Kularatna S, Byrnes J, Moyle W, Comans T. Developing a dementia-specific health state classification system for a new preference-based instrument AD-5D. Health & Quality Of Life Outcomes. 2017;15-21.
302.   King MT, Costa DS, Aaronson NK, Brazier J, Cella D, Fayers P et al. PCN98 The development of an internationally valid cancer-specific multi-attribute utility instrument (MAUI) from the Eortc Core Health-Related Quality of Life (HRQOL) Questionnaire, QLQ-C30. Value In Health. 2012; 15(4). A1–A256.
303.   Chen G, Ratcliffe J. A review of the development and application of generic multi-attribute utility instruments for paediatric populations. Pharmacoeconomics. 2015;33(10):1013-28.
304.   Horwood L, Li P, Mok E, Oskoui M, Shevell M, Constantin E. Health-related quality of life in Canadian children with cerebral palsy: what role does sleep play? Sleep Medicine. 2019;54:213-22.
305.   Chen K-L, Tseng M-H, Shieh J-Y, Lu L, Huang C-Y. Determinants of quality of life in children with cerebral palsy: A comprehensive biopsychosocial approach. Research in developmental disabilities. 2014;35(2):520-8.
306.   Angner E, Ray MN, Saag KG, Allison JJ. Health and happiness among older adults: a community-based study. Journal of Health Psychology. 2009;14(4):503-12.
307.   Diener E, Pressman SD, Hunter J, Delgadillo-Chase D. If, why, and when subjective well-being influences health, and future needed research. Applied psychology Health and well-being. 2017;9(2):133-67.
308.   Medvedev ON, Landhuis CE. Exploring constructs of well-being, happiness and quality of life. PeerJ. 2018;6:e4903. doi:10.7717/peerj.4903.
309.   Erdem S, Thompson C. Prioritising health service innovation investments using public preferences: a discrete choice experiment. BMC Health Services Research. 2014;14,360. https://doi.org/10.1186/1472-6963-14-360
310.   Costa D, King M, Aaronson N, Brazier J, Cella D, Fayers P, et al. The development of an internationally valid cancer-specific multi-attribute utility instrument (MAUI) from the EORTC Core HRQOL questionnaire, QLQ-C30. 2013; 8(S3); 335-33
311.   Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. Medical Decision Making. 2015;35(3):328-40.
312.   Mulhern B, Norman R, Viney R, Stolk E. Using discrete choice experiments to value generic preference-based measures: a systematic review. Value in Health. 2016;19(3):A95-A6.
313.   Brazier J, Deverill M, Green C. A Review of the use of health status measures in economic evaluation. Journal of Health Services Research & Policy. 1999;4(3):174-84.
314.   Ratcliffe J, Huynh E, Stevens K, Brazier J, Sawyer M, Flynn T. Nothing about us without us? a comparison of adolescent and adult health-state values for the Child Health Utility-9D using profile case Best-Worst Scaling. Health Economics. 2016;25(4):486-96.

315.   Tonmukayakul U, Imms C, Mihalopoulos C, Reddihough D, Carter R, Mulhern B, et al. Health-related quality of life and upper-limb impairment in children with cerebral palsy: developing a mapping algorithm. Developmental Medicine & Child Neurology. 2020;62(7):854-60.
316.   McDonough CM, Grove MR, Tosteson TD, Lurie JD, Hilibrand AS, Tosteson ANA. Comparison of EQ-5D, HUI, and SF-36-derived societal health state values among spine patient outcomes research trial (SPORT) participants. Quality Of Life Research. 2005;14(5):1321-32.
317.   Choi S, Choi JA, Kwon JW, Park SM, Jee D. Utility values for glaucoma patients in Korea. PLoS One. 2018;13(5):e0197581. doi:10.1371/journal.pone.0197581.
318.   Bowling A. Just one question: if one question works, why ask several? Journal of Epidemiology and Community Health. 2005;59(5):342-5.
319.   Bowling A, Bond M, Jenkinson C, Lamping DL. Short Form 36 (SF-36) Health Survey questionnaire: which normative data should be used? Comparisons between the norms provided by the Omnibus Survey in Britain, the Health Survey for England and the Oxford Healthy Life Survey. Journal of Public Health Medicine. 1999;21(3):255-70.
320.   Luyten J, Marais C, Hens N, De Schrijver K, Beutels P. Imputing QALYs from Single time point health state descriptions on the EQ-5D and the SF-6D: a comparison of methods for Hepatitis A patients. Value in Health. 2011;14(2):282-90.
321.   Topp J, Andrees V, Heesen C, Augustin M, Blome C. Recall of health-related quality of life: how does memory affect the SF-6D in patients with psoriasis or multiple sclerosis? A prospective observational study in Germany. BMJ Open. 2019;9(11):e032859. doi:10.1136/bmjopen-2019-032859.

Appendices

Appendix A Griffith ethical clearance

GRIFFITH UNIVERSITY HUMAN RESEARCH ETHICS COMMITTEE

Dear Prof Paul Scuffham

I write in relation to your application for ethical clearance for your project "A valuation study to develop a Cerebral Palsy specific preference-based measure" (GU Ref No: 2018/930). The research ethics reviewers resolved to grant your application a clearance status of "Fully Approved".

This is to confirm receipt of the remaining required information, assurances or amendments to this protocol.

Consequently, I reconfirm my earlier advice that you are authorised to immediately commence this research on this basis.

The standard conditions of approval attached to our previous correspondence about this protocol continue to apply.

Regards

Kim Madison | Human Research Ethics

Office for Research
Griffith University | Nathan | QLD 4111 | Level 0, Bray Centre
T +61 7 373 58043 | email k.madison@griffith.edu.au

Researchers are reminded that the Griffith University Code for the Responsible Conduct of Research provides guidance to researchers in areas such as conflict of interest, authorship, storage of data, & the training of

CHAPTER 2:

Appendix B Systematic review models information

Table A2.1.1 the general info of the models used in the systematic review

author	year	MAUI	Analyse model	Choice sets	DCE pairs each person	Design method N: non zero priors Z:zero priors	Population	Population represent general population by	Admin	software
Gamper et al(1)	2018	QLU-C10D	conditional logit model	960 choices sets	16 DCE choices	-	Germany and France General population	Age, sex and educational level	company that specializes in online DCE surveys	-
Selivanova et(2) al	2018	EQ-5D-5L and EQ-5D-3L	conditional logit model	240 choice sets for each survey	16 DCE choices	D-efficient (N)	Dutch (The Netherland) general population	Age and sex	Survey Sampling International (SSI)	Stata-Ngene
Cole et al(3)	2018	EQ-5D-5L	conditional logit model	196 choice sets (28 blocks)	14 DCE choices	Bayesian efficient design-N	UK general public	Age and gender.	EpiGenesys	Stata
Craig et al(4)	2018	EQ-5D-5L	Zermelo-Bradley-Terry (ZBT) Model	164,440 paired comparison	20 DCE choices	-	US respondents from all 50 states and Washington	Nationally representative panel	-	Ngene
Devlin et al(5)	2018	EQ-5D-5L	latent calss + hybrid DCE TTO	196 choice sets (28 blocks)	7 DCE choices	Bayesian efficient design-N	UK general population	Age	Research company Ipsos MORI	R and Winbugs

King et al(6)	2018	EORTC QLU-C10D	conditional logit model and mixed logit	960 choices sets	16 DCE choices	orthogonal main effect	Australian general population	Age and sex	SurveyEngine	Stata
Rowen et al(7)	2018	Diabetes Health Profile-Five Dimension	conditional logit model	120choice sets (10 versions)	12 DCE choices	D-optimal methods	UK general population	Age and sex.	market research agency	Ngene
Jonker et al(8)	2018	SF-6D	multinomial logit model	224 paired comparisons	14 choice tasks	Bayesian efficient design-N	The Netherlands general population	Aged 18 years and over	the Longitudinal Internet Studies for the Social Sciences (LISS)	Matlab
P.F.M Krabbe et el(9)	2017	EQ-5D-5L	multinomial probit model	200 choice sets (20 blocks)	10 choice sets	Bayesian efficient design-N	Canada, England, The Netherlands, and the United States	Age, education, and sex	the EuroQol Valuation Technology	Stata
Xie et al(10)	2017	EQ-5D-5L	conditional logistic regression model +Random effects	196 choice sets (28 blocks)	7 DCE choices	D-efficient (N)	UK, Canada, Spain, the Netherlands, China, Japan, Korea, and Thailand	Age and sex	commercial survey companies	
Purba et al(11)	2017	EQ-5D-5L	conditional logit model	196 choice sets (28 blocks)	7 DCE choices	Bayesian efficient design-N	Indonesian general population	Multi-stage stratified quota method with respect to residence, gender, age, level of education, religion and ethnicity	-	Stata
Huynh et al(12)	2017	ICECAP-SCM	multinomial logit model	second eight were from one of five blocks	16 DCE choices	Bayesian D-efficient-N	UK general population	Gender	-	Ngene

				within the design						
Mulhern et al(13)	2017	EQ-5D-5L	Conditional logistic regression model	120 choice sets	10 DCEtto pairs	D-efficiency-Z	UK general population	Age and Gender	commercial internet panel (IPSOS Observer)	Stata and Ngene
Ludwig et al(14)	2017	EQ-5D-5L	Not mentioned	196 choice sets (28 blocks)	7 DCE choices.	Efficient design	General population in 2 German cities	Not mentioned	computer assisted personal interview (CAPI) survey and EuroQol (EQ-VT)	Stata
Ramos-Gon et al(15)	2017	EQ-5D-5L	Conditional logit regression	196 choice sets (28 blocks)	7 DCE choices.	Bayesian design.-N	Spanish population	distribution of employment status; mean age and sex	EuroQol Valuation Technology	Stata
Mulhern et al(16)	2016	EQ-5D-5L	Conditional logistic regression model	120 choice sets	10 DCEtto choices	D-optimal	Australia general population	Age and Gender	Pure Profile	Stata
Mulhern et al(17)	2016	EQ-5D-5L	Conditional logistic regression model	42 choice sets (6 arm)	7 DCE choices	None	South York shire in the UK	Not mentioned	Not mentioned	Stata
Norman et al(18)	2016	QLU-C10 from Cancer Quality of Life (Core 30) (EORTC QLQ-C30)	Conditional logistic regression model	960 choice sets	16 DCE choices	Two-factor interactions	France and Germany	Not mentioned	internationa l Multi-Attribute Utility in Cancer (MAUCa) Consortium	Stata

Norman et al(19)	2016	QLU-C10D derived from the (EORTC QLQ-C30)	Conditional logistic regression model	960 choice sets	16 DCE choices	Orthogonal main effects plan.	Australia	male/female mix and proportion of Aboriginal /Torres Straits Islanders	SurveyEngine	Not mentioned
Rowen et al(20)	2016	EQ-5D-3L	Conditional logistic regression model	580 choice sets	10 DCE choices	D-optimality algorithm	UK general population	Age (min 18) and Sex	Not mentioned	Stata
Shiroiwa et al(21)	2016	EQ-5D-5L	Conditional logistic regression model	86 choice sets	10 DCE_{tto} choices + 1 death	Bayesian efficient-N	Japan general population (Tokyo, Okayama, Nagoya, Osaka, and Niigata	Age and Sex	Research company (ANTERIO Inc.)	SAS R WinBUGs
Scalone et al(22)	2015	EQ-5D-3L	Conditional logistic regression model	60 pairs of choice sets	9 DCE choices	Fractional factorial-Bayesian efficient-N	The Netherlands	University students	EuroQol Group	Stata
Eckert et al(23)	2015	EQ-5D-3L	Multilevel mixed effects logit regression analysis	243 choices	15 case vignettes used for DCE	Not mentioned	General German population from East German state	gender, age ,educational level	The authors	SAS
Rowen et al(24)	2015	asthma condition specific measure (AQL-5D)	Random effects probit model	24 choice sets (4 versions)	6 DCE choices	D-efficiency approach-Z	UK general population valuation	representative sample of the general population but no term mentioned	Not mentioned	SAS and R
Bansback et al(25)	2014	EQ-5D-5L	Conditional logistic regression model	120 choice sets	15 DCE_{tto} choices	D-efficient design using Fedorov algorithm- Z	UK general population	Age and Gender	The survey was designed and hosted by the market	Stata

								research company	
Krabbe et al(26)	2014	EQ-5D-5L	Multinomial probit regression model	200 choice sets	10 DCE choices	Bayesian algorithm-N	Canada, England, The Netherlands and US general population	Age, Education, and gender	the EuroQol Valuation Technology	Stata
Norman et al(27)	2014	SF-6D	Random-effects probit model.	180 choice sets (12 blocks)	15 DCE choices	D-efficient using 2-factor interactions-Z	Australian general population	Older age group is under represented and the individuals are better educated	Pure Profile Pty Ltd	SAS and Stata
Viney et al(28)	2014	EQ-5D-3L	Conditional logistic regression model	1620 choice sets (108 blocks)	15 DCE choices	Two-factor interaction	Australian general population	Gender and household income and less representative of age	Pure Profile Pty Ltd	Not mentioned
Norman et al(29)	2013	EQ-5D-5L	Random effects probit	200 choice sets (20 blocks)	10 DCE choices	Orthogonal main effects plan (OMEP)	Australian general population	Younger and better educated	Pure Profile Pty Ltd	Stata
Ramos -Goni et al(30)	2013	EQ-5D-5L	Conditional logistic regression model	50 choice sets (5 blocks)	10 DCE choices	Bayesian efficient-N	Spanish general population	Age, Gender and level of education	EuroQol Valuation Technology (EQ-VT) software	Stata
Prosser et al(31)	2013	influenza-related health-related quality	Generalized estimating equation and Random-effects probit regression	264 scenarios in the design for uncomplicated influenza and 136 scenarios for hospitalization	2 set of 8 DCE choices	Full factorial design	US adults randomly sampled from an Internet survey panel	Comparing different age groups	Knowledge Networks,	Stata

Pullenayegum et al(32)	2013	EQ-5D-5L.	Conditional probit regression through Generalized linear mixed models	400 choice sets (200pair with (20 blocks)	10 DCE choices	Bayesian efficient design by computing the D-error-Z	respondents from Canada and UK general population	Age and Gender	EuroQol Group Valuation Technology (EQ-VT)	Not mentioned
Bansback et al(33)	2012	EQ-5D-3L	Conditional logistic regression model	144 choice sets (24 blocks)	12 DCEtto choices	D-optimality algorithm	Canada General Population (English-speaking)	Age, Gender and level of education (min age 18)-	originally developed by the York Measurement and Valuation	SAS and Matlab
Ratcliff et al(34)	2012	CHU9D	Conditional logistic regression model	50 choice sets (5 blocks)	10 DCE choices	Fractional factorial design, near orthogonality	Australian (not representative adolescent population)	high socioeconomic status	Health (MVH) Group	Stata
Brazier et al(35)	2012	AQL-5D and The OAB-5D	Rank order logit model also Conditional logit model	24 choice sets (4 versions)	6 DCE choices	D-efficiency approach.	General public in South Yorkshire (UK)	Two samples same in the socio-demographic composition	the York Measurement and Valuation Health Group, which uses a 'time board' as a visual aid	SAS
Stolk et al(36)	2010	EQ-5D-3L	Conditional logistic regression model And Rank ordered logit model	60 choice sets	18 DCE pairs students 9 DCE pairs General	Bayesian efficient approach-N	Dutch general population and students(the Netherlands)	Age, Gender, level of education	administered to a community-based sample of consenting adolescents	Stata and Ngene

Ratcliffe et al(37)	2009	SQOL-3D	Random Effect probit model	64 choice sets	12 DCE choices	Huber and Zwerina design criteria	UK General population	Not mentioned	Not mentioned	SAS
Burr, et al(38)	2007	Glaucoma utility index	Conditional logistic regression model	32 choice sets	32 DCE choices	Fractional factorial design	Two ophthalmology centres UK (Aberdeen and Leeds). hospital-based glaucoma clinics	Male and female patients with different stages of disease, different ethnicity, and different ages	Not mentioned	Stata

CHAPTER 2

Appendix C DCE Designing methods

Table A2.1.2: methods on designing a DCE

author	year	Number of participants	Using other preference based measures except DCE	including Death as an option	randomising attributes	logic consistency	anchoring	scaling anchor	in addition to health states choice
Gamper et al	2018	N=605 Germany=300 and France=305	No	No	Blocks and ordering	Utility coefficients were checked	Bansback et al	Not mentioned	n
Selivanova et al	2018	N=4036	No	No	Blocks	removing dominant tasks and checking the coefficients	Worst health state- best health state	Best health state-WTD	n
Cole et al(3)	2018	N=993	No	No	Blocks and DCE arm	including two fixed tasks that contained a dominated option	-	-	n
Craig et al(4)	2018	N=8222	No	yes	Block	logical consistency of coefficients was examined for model performance	-		n
Devlin et al(5)	2018	N=996	TTO	no	Blocks , ordering	-	Worse than dead - best health state	Best health state-WTD	n
King et al(6)	2018	N=1846	DCEtto	no	Blocks , ordering of choice sets	-	-	Best health state-PITS (worse health state)	n
Rowen et al(7)	2018	n = 1493	DCEtto	no	Block	logical consistency of coefficients was	Coefficients were anchored using MRS	Dead - full health	n

						examined for model performance	by Dividing the interaction coefficients by the coefficient for duration- Bansback 2012		
Jonker et al(8)	2018	N=1775	DCEtto	no	Block	Having the correct sign and being statistically significant	Duration included - Bansback 2012		n
P.F.M Krabbe et el	2017	N=1775, 547 for Canada, 404 for the UK, 407 for the Netherlands, and 417 for the US	VAS	no	Blocks and order of the pairs and order within each pair	Coefficients were checked	Worst health state-best health state	Best health state-WTD	n
Xie et al(10)	2017	varied between 983-1299	cTTO	no	Blocks	-	-	-	n
Purba et al(11)	2017	N=1054	cTTO - VAS	Not mentioned	Blocks, pairs in random order,	coefficients are said to be logically consistent if magnitude values from logically worse health states are lower than those from logically better health states	Model 4: Combined DCE with C-TTO in a 'hybrid model', imposing the (0) death to (1) full health scale as determined by C-TTO.	Dead - full health	n
Huynh et al(12)	2017	N=6020	BWS	no	Blocks	-	This rescaling ensures that the 'no capability' state sums to zero, that 'full capability' state sum to one w	Best health state-WTD	n

								n
Mulhern et al(13)	2017	N=802	DCE-tto	Not mentioned	Blocking	Respondents who consistently chose the health profiles with the longest duration	Dividing the interaction coefficients by the coefficient for duration- Bansback 2012	Full health-worse than dead (WTD)
Ludwig et al(14)	2017	200 interviews	cTTO	yes	Blocks	A health state that dominates the other one.	values for the health states BTD range between 0 and 1)and values for health states WTD range between −1 and 0	Best health state-WTD
Ramos-Gon et al(15)	2017	N=973	cTTO	Not mentioned	blocks, ordering,	Coefficients of worse health states are lower than better health states	Model 3:The values obtained from DC models are expressed on an arbitrary scale and need to be rescaled on the dead (0) full health (1) scale	Full health - Dead
Mulhern et al(16)	2016	N=1080	Best and Worst of the scenarios	yes	Arm 2 randomised order between respondents, and arm 3 randomised within respondents.	Dimension ordering was used to check the consistency	Bansback et al 2012, the anchored values for each level of each dimension are produced by dividing the interaction coefficients by the coefficient for duration (both estimated from the conditional logit regression	Full health-Death
Mulhern et al(17)	2016	N=456	TTO for better than dead and worse than	yes	The health state dimensions and the positions of	Higher score for one state, A, than for another, B, if A was	method 3 mapping DCE on TTO	Full health - Death

Note: The rightmost column header is "n" (appearing in top rows). The last row's rightmost cell reads "immediate dead".

			dead health states		the states were randomized	logically better than B on at least one dimension and no worse on any other dimension.			
Norman et al(18)	2016	N =2053 (Germany=1002 and France =1051)	Not mentioned	Not mentioned	Ordering	Not mentioned	Bansback et al 2012	Not mentioned	n
Norman et al(19)	2016	N=430	Not mentioned	Not mentioned	Blocks, ordering .	Not mentioned	Bansback et al 2012	Not mentioned	n
Rowen et al(20)	2016	N=3669	added life expectancy in their DCE	not in the questionnaire	Blocks	Not mentioned	The marginal rate of substitution (MRS) was used and it was estimated using the ratio of the marginal utilities		n
Shiroiwa et al(21)	2016	194 =Tokyo, 147 = Okayama, 210 = Nagoya, 235 = Osaka, 240=Niigata N= 1026	TTO	-	Block.	Not mentioned	All models	Full health - WTD	n
Scalone et al(22)	2015	N=208	DCEtto (up to 50 years)	Not mentioned	Bounded randomization procedure,	check for the presence of systematic effects on choices (chooses the same left health profile every time)	Bansback et al 2012	Full health - dead	n
Eckert et al(23)	2015	N=683	VAS	Had both or neither in the choices	Case vignettes were assigned randomly to the pairwise	Administered two questionnaire versions with different ordering of questions to test for	Not mentioned	Best imaginable health- Worst imaginable health	4options

					choice scenarios and questionnaires	possible priming effects.			
Rowen et al(24)	2015	n = 263	TTO	Not mentioned	Respondents were selected randomly	-	All models	Full health-dead	n
Bansback et al(25)	2014	N= 1799	DCEtto	Not mentioned	The scenarios were randomly allocated	Coefficients for each attribute were compared to identify if worse levels had lower values	Model 3: the values for each individual health state can be anchored using the estimated coefficients	Not mentioned	n
Krabbe et al(26)	2014	547 =Canada, 404 = England, 407= The Netherlands, and 417 =the US,	No	No	ordering	Logical ordering of parameter estimates	Undefined scale (without meaningful anchors)	Full health and death	n
Norman et al(27)	2014	N = 1004	Best and worse	yes	Respondents were randomly assigned to a choice task	Not mentioned	Bansback et al 2012	Full health and death	immediate dead
Viney et al(28)	2014	N= 1031	Best and worse + death as an option	yes	Respondents were randomly assigned to choice sets.	Not mentioned-	Bansback 2012- impose the QALY constraints on the regression and then take the ratio of marginal utilities, an approach analogous to that of	Full health and death	immediate dead
Norman et al(29)	2013	N= 930	Best and worse + death as an option	yes	Not mentioned	pairs in which the respondent picked the option that provided shorter duration and poorer	Bansback et al 2012	Full health-WTD	immediate dead

						quality of life (which we have termed, for convenience, the dominated option)			
Ramos -Goni et al(30)	2013	N= 400	DCE-tto and DCE-wtd	DCE-dead	Not mentioned	logical ordering and omitted the participants without logical order	Model 2: the worst health state predicted on the lead-time TTO model was taken as an anchor point to rescale the arbitrary scale of the conditional logistic mode	Dead-full health	n
Prosser et al(31)	2013	N = 1,012	TTO and WTP	Not mentioned	Scenarios were then randomly blocked into groups of eight questions.	One dominated choice was retained (at random) in each block	Not mentioned	Not mentioned	n
Pullenayegum et al(32)	2013	N= 545 Canada and N = 403 UK	TTO	Not mentioned	Random call was the primary for recruitment. Also the health states were chosen randomly	Not mentioned	Health utilities are anchored at 0 for death and 1 for full health and have interval properties	Full health – dead	n
Bansback et al(33)	2012	N= 693	TTO and comparing it with DCEtto	Not mentioned	Randomised for each respondent	Two tasks were included, where one of the two profiles was regarded as a dominant option	Bansback et al 2012 mean DCETTO value can be calculated from the coefficients of the conditional logit mode	Full health – dead	n
Ratcliff et al(34)	2012	N = 590	BWS instead the traditional	Not mentioned	Not mentioned	Not mentioned	Mapping on SG method result	Best health state-PITS	n

			DCE (BWS DCE)					(worst health state)	
Brazier et al(35)	2012	N=263 for AQL-5D and N=402 OAB-5D	TTO and Ranking order	yes	blocks	Only one pairwise comparison involves a logically consistent choice where one state has better health for every dimension.	Model 1 and 2: normalising using (1) the TTO value of the worst state and (2) the coefficient on the 'dead' dummy variable.	Full health- dead	n
Stolk et al(36)	2010	general population sample N = 438 and student sample N = 204	used TTO and VAS ranking	DCEdead	Health state were randomly selected from the choice set. Alterations were made randomly.	Identified and altered dominant choices in which logical consistency predicts that one alternative will always be preferred.	Model 1: The value for being dead is anchored at zero by dividing all coefficients by the coefficient for "dead."	dead- full health	n
Ratcliffe et al(37)	2009	N=102	TTO and Ranking order	Not mentioned	Choices were randomly distributed between the two versions of the questionnaire	Had logical consistencies in coefficients.	Model 3: mapping DCE on TTO	perfect health- PITS	n
Burr, et al(38)	2007	N=289	no	Not mentioned	Not mentioned	This study had a consistency in respondent choice.	Best health calculated by summation of the coefficients associated with the best level for each dimension. The weights for all other levels of each	Perfect health - death	n

							dimension were then estimated as a proportion of this score, allowing all combinations to be estimated on a 0 to 1 scale		

CHAPTER 2:

Appendix D Search terms

Data bases:

Medline (Ovid MEDLINE 1946 to September 2018), EMBASE, Cochrane Library and CINAHL (Ebsco).

Word search:

1. discrete choice experiment.mp.

2. exp Patient Preference/

3. discrete choice experiment$.mp.

4. (discrete choice model$ or stated preferenc$ or paired comparison& or pairwise choices).mp.

5. 1 or 2 or 3 or 4

6. exp Quality-Adjusted Life Years/

7. quality adjusted life year$.mp.

8. QALY.mp.

9. quality of life.mp.

10. EQ 5D.mp.

11. Child Health Utility instrument.mp.

12. CHU9D.mp.

13. AQol.mp.

14. SF-36.mp.

15. Short Form Health Survey.mp.

16. SF-6D.mp.

17. SF-12.mp.

18. HRQL.mp.

19. Health Utilities Index.mp.

20. HUI.mp.

21. Sickness Impact Profile.mp.

22. SIP.mp.

23. 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22

24. 5 and 23

CHAPTER 3

Appendix E Detailed explanation of Rasch and Factor analysis:

Factor and Rasch Methods: There are two types of factor analysis. Exploratory factor analysis (EFA), which is based on data; this model is used to discover the structure of the factors and has criteria for reducing the factors. EFA can be used to evaluate the construct validity of a test, instrument or scale and addresses multicollinearity of two or more variables in the model. EFA is used for the development of parsimonious interpretation, analysis, and theoretical constructs. Confirmatory factor analysis (CFA), which is based on theory; this model is applied when confirming a hypothesised factor structure (39).

Factor analysis can use orthogonal or oblique rotations. Oblique rotation (such as Promax) is used when the factors are correlated and is often produces more accurate results when human behaviour is involved. In contrast, the orthogonal rotation (such as Varimax) is employed when the factors are uncorrelated (40).

To show item difficulty in Rasch, the items are scaled based on their difficulty, where they are located on a measurement scale, in which easier items are located on the left and moving toward the right, items get more complicated. The person measure is a quantitative measure of a person's attitude on a uni-dimensional scale. In Rasch analysis, the term "Person measure" is the Rasch scale number that shows the functioning of a test taker or scale respondent.

Rasch analysis belongs to a class of Item response theory (IRT) models. Compared to classical test theory (CTT), IRT is a modern approach to test the reliability of psychological test assessment, assessing how and why a person responds to an item(41). Both methodologies can be used to assess individual change in the clinical context. However, when estimating measurement precision, they differ. CTT assumes that it will be equal for all participants regardless of their attribute level, but IRT assumes that it depends on a latent-attribute (unobservable) value. This difference leads results with different conclusions regarding the change in statistical significance (42).

IRT contains models for dichotomous and polytomous responses. For dichotomous response data, there are one parameter, two parameters or three parameters logistic models. Models for polytomous data include partial credit model (PCM), rating scale model (RSM), generalised partial credit model (GPCM), graded response model (GRM), nominal response model (NRM). PCM and RSM assume that the discrimination parameter is equal across all items and as such, belong to the Rasch family.

CHAPTER 3:

Appendix F Item statistics

Supplementary Table S 3.1: Items summary statistics

		N (473)	Missing (%)	Mean	Std. Dev.	% response at floor (L1)	% response at ceiling (L9)	Loadings
social wellbeing and acceptance	get along with other children outside of school	471	0.42	7.34	1.37	0.21	21.87	0.724
	get along with adults	471	0.42	7.62	1.26	0.21	26.96	0.796
	accepted by other children outside of school	470	0.64	7.16	1.47	0.43	19.15	0.674
	accepted by adults	472	0.21	7.61	1.34	0.21	31.57	0.774
	accepted by people in general	468	1.07	7.44	1.32	0.21	24.36	0.707
physical health	ability to participate at school	466	1.50	6.95	1.53	0.86	16.31	0.481
	ability to participate in recreational activities	472	0.21	6.70	1.72	0.64	15.89	0.510
	participate in sporting activities	470	0.64	6.33	1.83	0.64	12.55	0.500
	their physical health	472	0.21	6.99	1.66	0.21	19.92	0.463
	the way they get around	473	0.00	6.81	1.67	0.21	16.49	0.573

	the way they look	473	0.00	7.22	1.58	0.42	25.58	0.585
	their ability to keep up physically with their peers	472	0.21	5.64	1.91	2.75	6.78	0.777
	feels about themselves	471	0.42	7.32	1.48	0.42	25.27	0.558
	the way they use their legs	471	0.42	6.08	1.80	1.27	10.19	0.697
communication	their ability to participate in social events outside of school	470	0.64	7.01	1.58	0.43	19.79	0.727
	their ability to participate in your community	469	0.85	6.94	1.52	0.21	16.84	0.746
	the way they communicate with people they know well	472	0.21	7.81	1.26	0.21	38.56	0.484
	the way they communicate with people they don't know well	473	0.00	6.76	1.65	0.42	15.64	0.769
	the way people communicate with them	472	0.21	7.21	1.38	0.42	20.76	0.713
Pain and discomfort	How much pain do your child have	470	0.64	2.53	1.82	0.64	39.15	0.925
	the amount of pain they have	449	5.35	2.71	2.09	1.78	39.87	0.904
	How much discomfort does your child experience	469	0.85	2.91	1.83	0.43	25.37	0.836

CHAPTER 3:

Appendix G CPQOL levels

Supplementary file 3.2:

-Levels in the original instrument:

Very unhappy	unhappy	Neither Happy nor unhappy	happy	Very happy

1	2	3	4	5	6	7	8	9

-How questions present items:

Question: How do you think your child feels about *how they are accepted by people in general?*

Item: *accepted by people in general*

Appendix H Ngene codes for the DCE choice tasks design in valuation approach DCETTO

- *DCE TTO with dummy interactions and including duration*
- *Duration attribute: 1 year, 3 years, 5 years , 7 years and 10 years*
- *D-efficient (with priors from the pilot study)*

```
;alts = alt1, alt2
;rows = 240
;eff = (mnl,d)
;block = 20
;model:
U(alt1)=
b1[0.394] * DUR[1,3,5,7,10](38-58,38-58,38-58,38-58,38-58) +
b2.dummy[0|0|0|0] * Accpt[1,2,3,4,0] +
b3.dummy[0|0|0|0] * Phys[1,2,3,4,0] +
b4.dummy[0|0|0|0] * Mnul[1,2,3,4,0] +
b5.dummy[0|0|0|0] * Comm[1,2,3,4,0] +
b6.dummy[0|0|0|0] * Pain[1,2,3,4,0] +
b7.dummy[0|0|0|0] * Slp[1,2,3,4,0] +

i1[-0.002] * DUR * Accpt.dummy[1] +
i2[-0.068] * DUR * Accpt.dummy[2] +
i3[ -0.092] * DUR * Accpt.dummy[3] +
i4[-0.122] * DUR * Accpt.dummy[4] +
i5[0] * DUR * Phys.dummy[1] +
i6[-0.050] * DUR * Phys.dummy[2] +
i7[ -0.061] * DUR * Phys.dummy[3] +
i8[ -0.088] * DUR * Phys.dummy[4] +
i9[ -0.034] * DUR * Mnul.dummy[1] +
i10[ -0.061] * DUR * Mnul.dummy[2] +
i11[ -0.101] * DUR * Mnul.dummy[3]+
i12[ -0.144] * DUR * Mnul.dummy[4] +
i13[0] * DUR * Comm.dummy[1] +
i14[ -0.011] * DUR * Comm.dummy[2] +
i15[-0.070] * DUR * Comm.dummy[3] +
i16[ -0.071] * DUR * Comm.dummy[4] +
i17[0] * DUR * Pain.dummy[1] +
i18[-0.009] * DUR * Pain.dummy[2] +
i19[-0.026] * DUR * Pain.dummy[3]+
i20[-0.077] * DUR * Pain.dummy[4] +
i21[ -0.014] * DUR * Slp.dummy[1] +
i22[ -0.023] * DUR * Slp.dummy[2] +
i23[ -0.087] * DUR * Slp.dummy[3]+
i24[ -0.103] * DUR * Slp.dummy[4]
/
U(alt2)=
b1[0.394] * DUR[1,3,5,7,10](38-58,38-58,38-58,38-58,38-58) +
b2.dummy[0|0|0|0] * Accpt[1,2,3,4,0] +
b3.dummy[0|0|0|0] * Phys[1,2,3,4,0] +
```

```
b4.dummy[0|0|0|0]  * Mnul[1,2,3,4,0] +
b5.dummy[0|0|0|0]  * Comm[1,2,3,4,0] +
b6.dummy[0|0|0|0]  * Pain[1,2,3,4,0] +
b7.dummy[0|0|0|0]  * Slp[1,2,3,4,0] +

i1[-0.002]  * DUR * Accpt.dummy[1] +
i2[-0.068]  * DUR * Accpt.dummy[2] +
i3[ -0.092]  * DUR * Accpt.dummy[3] +
i4[-0.122]  * DUR * Accpt.dummy[4] +
i5[0]  * DUR * Phys.dummy[1] +
i6[-0.050]  * DUR * Phys.dummy[2] +
i7[ -0.061]  * DUR * Phys.dummy[3] +
i8[ -0.088]  * DUR * Phys.dummy[4] +
i9[ -0.034]  * DUR * Mnul.dummy[1] +
i10[ -0.061]  * DUR * Mnul.dummy[2] +
i11[ -0.101]  * DUR * Mnul.dummy[3]+
i12[ -0.144]  * DUR * Mnul.dummy[4] +
i13[0]  * DUR * Comm.dummy[1] +
i14[ -0.011]  * DUR * Comm.dummy[2] +
i15[-0.070]  * DUR * Comm.dummy[3] +
i16[ -0.071]  * DUR * Comm.dummy[4] +
i17[0]  * DUR * Pain.dummy[1] +
i18[-0.009]  * DUR * Pain.dummy[2] +
i19[-0.026]  * DUR * Pain.dummy[3]+
i20[-0.077]  * DUR * Pain.dummy[4] +
i21[ -0.014]  * DUR * Slp.dummy[1] +
i22[ -0.023]  * DUR * Slp.dummy[2] +
i23[ -0.087]  * DUR * Slp.dummy[3]+
i24[ -0.103]  * DUR * Slp.dummy[4]

$
```

CHAPTER 5:

Appendix I Ngene design matrix of the DCE choice tasks design with priors in valuation approach DCETTO

D error   0.0165

Choice situation	alt2. accpt	alt2. phys	alt2. mnul	alt2. comm	alt2. pain	alt2. slp	alt3. accpt	alt3. phys	alt3. mnul	alt3. comm	alt3. pain	alt3. slp
1	3	5	4	5	3	3	1	2	3	4	2	5
2	1	2	4	2	2	5	3	3	5	4	4	3
3	3	5	4	2	4	3	5	4	1	3	2	1
4	5	2	1	5	5	4	1	4	5	1	3	3
5	1	3	4	3	2	4	5	1	1	1	5	2
6	1	5	5	2	2	3	2	3	4	1	1	5
7	3	2	5	5	5	5	5	1	4	4	4	1
8	3	5	2	2	4	2	4	2	4	3	1	1
9	2	5	3	5	1	2	1	3	4	3	2	3
10	2	4	4	2	1	5	1	2	1	3	2	1
11	4	4	5	4	4	5	2	5	3	5	1	4
12	1	1	5	3	4	4	3	4	1	5	5	3
13	2	2	1	1	2	2	1	3	5	5	5	1
14	3	2	1	3	4	2	5	1	4	1	1	4
15	1	1	2	5	2	2	5	2	1	1	5	3
16	5	3	4	5	1	2	4	5	2	3	3	5
17	2	5	3	3	5	4	3	2	2	5	2	5
18	1	2	4	1	5	4	2	3	5	4	3	2
19	3	1	5	3	2	3	4	5	4	5	3	2
20	4	1	4	2	3	5	2	2	5	4	2	1
21	1	3	5	2	2	4	2	5	1	3	1	2
22	2	4	1	5	3	4	4	5	4	2	1	2
23	1	3	3	5	1	3	5	2	4	3	3	1
24	5	2	1	4	1	5	2	1	3	5	4	1
25	5	5	5	1	5	2	5	5	5	1	5	2
26	1	3	4	1	1	5	4	4	3	5	2	1
27	2	5	3	2	5	2	3	2	4	3	1	4
28	5	5	1	3	2	3	3	4	5	2	1	1
29	4	5	1	5	4	5	3	4	5	1	3	2
30	5	4	1	5	5	3	5	4	1	5	5	3
31	3	2	4	5	5	2	2	5	5	3	4	4
32	1	5	2	1	2	5	2	2	1	2	3	4
33	5	3	2	5	2	3	3	1	3	1	4	2
34	4	3	3	4	3	5	5	1	1	3	5	4
35	4	1	4	5	1	5	2	4	5	4	4	1
36	2	2	3	1	5	2	1	4	4	2	4	3

37	4	3	1	1	2	4	5	4	2	3	4	3
38	2	1	2	3	3	2	3	4	5	1	4	1
39	5	3	3	1	1	5	4	2	1	4	5	4
40	2	4	3	3	3	5	1	3	1	2	5	3
41	3	5	2	3	1	3	2	3	4	1	5	2
42	5	5	2	4	1	4	4	4	3	2	4	5
43	3	1	5	4	3	1	3	1	1	4	3	1
44	4	5	1	4	4	1	1	4	4	3	2	3
45	1	1	4	4	2	4	2	1	2	1	1	3
46	1	3	2	1	4	5	2	5	5	2	3	4
47	1	4	1	4	5	3	3	3	2	2	4	4
48	5	4	3	1	3	1	1	2	5	4	4	2
49	5	3	1	4	4	5	5	3	1	4	4	5
50	3	3	2	3	1	2	5	2	5	1	3	5
51	1	4	1	5	1	4	3	5	2	3	4	3
52	2	3	4	2	2	3	1	4	5	3	4	2
53	2	5	1	3	1	5	1	2	3	2	5	1
54	2	1	5	5	5	1	4	5	2	1	2	4
55	2	2	2	4	2	5	4	3	4	3	5	4
56	3	4	5	4	3	5	5	5	2	3	5	3
57	1	2	4	5	4	4	4	3	3	4	1	1
58	2	2	2	2	1	5	4	1	1	4	2	1
59	5	2	3	3	5	4	3	5	5	5	3	2
60	2	1	1	2	3	1	5	5	2	4	2	3
61	5	1	4	1	4	3	1	4	2	2	5	5
62	3	2	2	5	3	4	2	4	1	3	1	3
63	4	5	2	5	5	2	1	4	3	4	1	5
64	4	2	4	1	1	3	3	1	1	2	2	5
65	4	2	3	2	2	1	3	4	2	1	4	4
66	1	2	3	5	3	5	4	3	2	2	1	1
67	4	1	5	3	1	1	2	5	2	1	3	5
68	4	4	4	5	4	3	5	2	2	1	5	1
69	4	5	3	4	5	1	5	4	5	5	2	2
70	5	3	2	2	4	5	3	4	5	5	1	2
71	5	1	5	2	1	2	1	5	4	1	5	3
72	4	1	2	4	1	2	3	5	1	1	5	1
73	3	1	4	1	3	2	4	2	3	3	4	5
74	5	5	2	4	4	3	2	1	5	2	1	5
75	5	4	4	1	1	4	1	3	1	2	4	3
76	1	5	5	2	4	4	3	3	4	1	3	2
77	5	1	5	2	3	3	2	2	4	5	2	2
78	4	4	4	4	1	1	5	3	3	5	2	4
79	3	4	2	1	4	3	1	5	3	3	3	1
80	4	1	3	3	5	5	1	5	5	2	2	4
81	2	4	1	4	4	4	5	3	2	2	3	3
82	2	4	4	2	5	1	4	2	5	1	1	2
83	5	4	2	5	3	4	4	5	4	2	5	3
84	2	5	3	5	1	3	3	3	4	4	3	5

85	3	1	4	3	4	1	4	2	5	2	2	3
86	2	1	3	4	2	4	1	4	1	5	5	1
87	4	4	4	3	4	2	3	2	2	1	4	1
88	2	1	5	4	4	1	1	3	3	5	3	3
89	4	3	5	5	3	2	5	5	4	4	4	1
90	5	2	1	5	5	4	3	2	1	5	1	4
91	5	2	3	5	3	3	4	3	1	1	5	5
92	5	2	5	1	1	1	4	4	3	2	4	2
93	5	4	3	1	4	5	4	1	1	2	1	4
94	1	2	4	5	2	1	4	5	1	3	5	4
95	3	1	3	2	2	4	5	4	1	3	3	1
96	2	5	4	5	1	1	1	2	3	1	5	4
97	4	3	1	3	5	2	3	5	2	2	1	1
98	4	2	2	2	3	2	4	2	2	4	3	5
99	4	5	2	4	4	3	1	3	3	1	1	4
100	1	1	1	5	4	1	2	3	2	4	1	4
101	4	2	3	1	3	3	5	1	2	3	1	4
102	2	3	5	1	2	1	3	5	1	3	3	4
103	2	4	1	4	3	3	4	2	2	1	4	4
104	2	3	5	3	2	2	5	1	2	5	5	3
105	2	5	4	4	3	4	4	4	3	5	1	2
106	2	3	4	4	2	4	3	1	5	3	3	2
107	2	4	5	1	5	1	5	3	1	2	4	5
108	3	5	2	3	3	5	4	3	3	2	4	3
109	5	4	3	4	2	2	2	1	5	5	5	1
110	4	3	1	3	3	2	1	1	3	5	1	1
111	4	5	1	3	2	5	5	1	3	5	4	3
112	2	2	3	3	2	1	1	1	2	1	5	2
113	4	5	1	5	1	3	2	4	4	4	2	4
114	5	4	5	2	3	4	5	4	5	2	3	4
115	1	2	2	3	1	3	5	3	4	5	2	1
116	2	2	2	1	4	3	3	5	1	2	1	2
117	3	5	5	3	1	3	2	4	2	1	3	5
118	5	1	5	5	2	3	4	4	3	2	4	4
119	1	4	2	2	5	2	2	2	1	1	4	5
120	2	2	3	3	2	5	3	2	3	4	2	5
121	5	2	2	1	4	1	3	1	4	3	5	2
122	1	4	2	3	5	5	4	2	5	4	2	2
123	2	1	5	2	5	3	4	5	3	5	4	1
124	4	1	5	5	1	5	5	4	4	3	3	3
125	3	1	2	5	1	4	5	3	3	1	2	1
126	2	5	4	1	3	3	5	5	4	1	3	3
127	5	4	1	2	4	4	2	2	5	1	2	5
128	1	3	2	4	3	2	5	2	4	2	4	5
129	3	4	1	1	4	2	2	1	2	4	3	5
130	1	5	3	1	3	1	4	4	5	3	2	4
131	1	5	3	2	5	2	3	3	2	5	3	3
132	3	3	4	1	2	5	5	2	3	4	4	2

133	1	4	2	3	5	1	2	3	1	4	3	3
134	3	1	4	4	2	3	4	2	2	1	4	5
135	5	3	2	2	1	1	1	2	4	5	4	3
136	2	3	1	2	1	3	4	1	4	4	5	2
137	1	5	3	3	2	4	4	4	2	5	1	5
138	1	5	1	2	5	5	2	3	4	3	3	3
139	4	1	4	5	3	4	2	5	5	3	1	3
140	2	3	2	4	5	5	5	2	4	5	2	2
141	1	1	2	4	5	3	2	5	4	5	1	2
142	3	4	2	5	2	1	1	1	5	1	1	3
143	2	2	4	3	3	2	3	1	1	1	2	5
144	3	2	5	4	3	4	2	3	3	1	5	2
145	3	4	3	2	2	1	2	1	5	1	1	2
146	2	3	5	5	5	4	5	4	1	2	1	5
147	5	2	5	2	1	5	4	3	1	5	3	4
148	4	1	5	3	5	1	2	5	4	2	2	5
149	3	4	2	3	1	3	1	1	4	2	3	1
150	4	4	2	1	2	1	3	3	1	4	4	4
151	4	1	4	1	1	3	2	4	1	5	5	4
152	3	5	4	2	5	2	3	1	4	2	5	2
153	4	3	2	4	3	2	5	1	3	2	1	5
154	5	1	2	2	4	1	3	4	3	3	2	5
155	2	5	4	3	3	5	3	2	1	1	1	2
156	3	5	2	3	5	1	1	1	3	4	4	2
157	3	2	3	4	1	3	1	5	5	2	5	1
158	3	1	3	1	2	4	1	3	2	5	5	1
159	3	4	3	4	3	2	2	3	5	5	2	4
160	5	3	1	4	2	1	2	1	3	2	4	4
161	4	4	5	1	2	2	3	3	2	4	3	1
162	4	3	2	5	5	4	4	3	2	5	5	4
163	1	5	1	5	3	5	2	1	5	3	2	1
164	5	1	1	1	3	4	4	4	4	2	4	1
165	5	1	3	1	1	4	1	1	3	2	5	1
166	3	3	3	2	4	4	1	5	1	3	2	5
167	3	4	4	2	1	5	4	3	3	5	2	1
168	4	3	5	1	1	3	4	5	5	1	1	3
169	5	4	3	5	3	4	3	1	4	3	2	5
170	1	5	5	5	1	2	5	1	3	4	3	5
171	2	1	3	5	5	5	1	2	4	1	4	4
172	1	1	3	5	4	3	2	5	1	2	2	1
173	2	3	5	1	5	3	1	4	2	4	2	2
174	3	3	2	3	4	5	2	5	3	5	1	4
175	2	4	4	3	4	1	1	1	5	4	3	5
176	3	4	4	4	5	1	3	4	4	4	5	1
177	5	2	5	3	4	2	3	4	3	4	5	4
178	5	2	1	3	4	2	2	4	2	4	5	3
179	5	5	2	3	4	4	5	5	2	3	4	4
180	3	2	1	4	5	2	2	1	2	5	4	5

183

181	1	4	1	1	2	3	3	2	3	3	5	1
182	5	5	5	4	5	5	5	2	5	4	4	5
183	2	3	2	1	4	4	5	2	5	3	3	2
184	5	4	5	3	2	2	3	5	2	5	3	1
185	4	2	1	1	5	2	1	5	4	4	1	3
186	3	5	3	1	2	2	2	4	1	3	1	5
187	3	3	3	3	2	3	4	2	2	2	3	2
188	4	1	2	3	3	4	3	5	4	1	5	2
189	1	3	3	2	1	3	3	4	2	3	2	2
190	3	2	5	2	2	5	1	3	1	4	1	3
191	1	4	5	3	5	5	1	4	5	3	5	5
192	3	2	4	2	5	3	5	3	1	5	2	5
193	2	4	1	3	4	4	4	2	2	4	1	1
194	4	2	5	4	1	5	5	1	2	5	5	2
195	3	3	3	3	3	1	2	1	4	4	5	5
196	5	3	5	5	1	1	2	5	3	1	3	5
197	1	4	4	4	1	1	5	3	5	2	3	3
198	1	1	4	5	2	3	5	3	4	4	1	2
199	2	5	5	2	3	2	3	1	1	4	5	4
200	4	4	5	1	3	1	5	2	2	3	5	2
201	2	2	2	4	5	1	5	5	3	1	1	5
202	2	2	2	5	4	5	5	5	4	1	2	4
203	2	2	1	5	5	3	1	5	5	1	1	4
204	3	5	1	1	4	2	4	1	5	2	2	3
205	3	3	1	1	3	4	2	1	3	5	2	5
206	3	1	3	4	4	2	5	2	4	2	2	1
207	1	3	4	4	4	2	4	4	5	2	3	3
208	4	5	3	5	4	1	4	5	3	5	3	1
209	4	1	1	2	2	5	2	2	5	4	5	3
210	1	1	3	1	3	1	2	3	1	2	2	2
211	5	1	1	2	4	2	1	5	3	3	3	4
212	2	3	5	4	4	1	4	4	1	2	3	3
213	4	5	1	4	3	4	4	5	1	4	3	4
214	1	3	4	2	3	4	3	5	3	5	2	5
215	4	4	2	4	2	4	3	3	4	3	3	5
216	4	1	3	2	5	3	1	2	4	5	3	1
217	3	2	3	2	5	3	1	3	2	5	4	1
218	1	3	5	1	4	4	5	1	4	4	3	3
219	1	4	5	1	4	1	4	1	2	3	2	4
220	5	4	2	4	1	2	3	3	1	5	4	5
221	4	2	4	2	1	1	1	5	5	4	4	2
222	4	4	1	4	1	4	3	3	5	2	3	2
223	1	1	5	3	3	5	1	4	5	3	3	5
224	1	5	4	2	2	3	2	1	3	1	4	4
225	1	5	1	4	2	1	4	2	5	1	5	3
226	1	2	4	4	5	2	3	4	2	1	2	4
227	5	1	1	1	5	1	2	2	2	3	4	3
228	3	3	3	2	5	2	1	1	1	5	1	1

229	5	4	3	5	2	2	2	4	3	5	2	2
230	1	3	1	2	1	5	4	1	3	3	4	3
231	2	2	1	2	3	4	1	3	4	3	1	2
232	5	5	2	5	5	5	3	4	1	4	1	3
233	3	1	1	1	2	5	4	2	3	2	1	4
234	4	5	4	1	4	1	1	3	3	4	5	4
235	3	2	1	4	4	3	1	1	2	3	5	5
236	5	2	4	2	3	5	4	5	2	1	2	3
237	5	3	5	3	3	1	3	2	4	5	1	4
238	4	3	3	3	1	1	1	2	1	2	2	3
239	1	3	3	3	1	1	2	1	2	2	4	2
240	4	1	5	1	2	4	5	3	3	4	1	2

alt1- health state 1;alt 2- health state 1; dur-duration; accpt–acceptance; phys–physical healthe;mnul–manual
ability; comm–communication; pain–pain/discomfort; slp–sleep.
Duration levels codes (1-1 year, 2- 3years, 3- 5 years, 4- 7 years ,5- 10 years)
Dimension level codes (1-level 1, 2-level 2, 3-level 3, 4-level 4, 5-level 5)

185

Appendix J The online survey

A valuation study to develop a Cerebral Palsy specific preference-based measure (Griffith University reference number 2018/913)

Why is the research being conducted?

You are invited to participate in the research project that is explained below. If you do not wish to participate, it is okay to say no.

Cerebral Palsy is the most common physical disability during childhood. This study will help to develop health state values for a Cerebral Palsy specific quality of life instrument, which is called a multi attribute utility instrument. This instrument will help to identify the most important factors effecting children with Cerebral Palsy. The generated values will let researchers and policy-makers make better decisions about the effectiveness of new interventions and treatments.

This research will contribute to Miss Bahrampour's studies for the award of the Doctor of Philosophy (PhD) degree course being undertaken at Griffith University.

What you will be asked to do

You will be asked to complete a short, online choice task, the first page is an introduction to the research and you will be asked to provide consent in order to continue with the rest of the survey. If you accept to be a part of this study, the next section will require you to provide demographic data about your child which includes your child's age, gender, and education. In addition to the demographic questions, you will also be asked about your child's gross motor functioning and the manual ability. Next, you will be required to do a choice experiment, in which you are asked to choose your child's most preferred health state (you think your child will choose) between two different health-states.

The expected benefits of the research

The values developed in this study will help inform policy makers. In this area, the measurement of utility required for economic evaluation is absent. This instrument will help researchers and assist health care funding agencies in Australia in assessing the effectiveness of newly developed health care. Our aim is to provide a more precise and accurate measurement of changes associated with quality of life than is currently undertaken. There is no expected direct benefit of this research to you.

Risks to you

There are no anticipated risks to you or your child from participating in this study. Answering all questions in the survey would take approximately 15 to 20 minutes of your time. The questionnaire will be completed online. You may complete the questionnaire at your convenience. It is important for us that you do not disrupt your normal activities because of our research.
The questionnaires will ask questions about your child's physical symptoms and psychological wellbeing. If you have any health concerns arising from the questionnaires contact your general practitioner. The questionnaires will ask you ask questions about your child's self-perception and mental wellbeing. If you or your child experience a personal crisis, help is available from Lifeline by calling 13 11 14, or through their website at www.lifeline.org.au.

Your confidentiality

We do not ask personal identification details in this survey. Your data will be treated confidentially at all times and your identity will not be revealed in any publication reporting the findings of this research study. Your anonymous survey responses may be used in future studies similar to this one. Research team will not have access to your personal identifications information at any stage of the research. All research data (survey responses and analysis) will be retained in a password protected electronic file at Griffith University for a period of five years before being destroyed.

Your participation is voluntary

Your participation in this research activity is absolutely voluntary. You are free not to participate. Even after you give consent to participate, at any moment during the survey you can withdraw your consent and your information at any stage of the data collection or after you have completed the survey. You will not be penalized or affected in any way for participating (or not participating) in this research, and your access to health care will not be affected in any way.

Questions / further information

If you wish to obtain more information on this research you can contact Miss Mina Bahrampour on her email: mina.bahrampour@griffithuni.edu.au or phone: (07) 3735 9110. If you wish to obtain a copy of summary results please contact Professor Paul Scuffham email: p.scuffham@griffith.edu.au or phone: (07) 3735 9132.

The ethical conduct of this research

This study is being conducted by Griffith University in accordance with the *National Statement on Ethical Conduct in Human Research* (2007). If you have any concerns or complaints about the ethical conduct of the research project they should contact the Manager, Research Ethics on (07) 3735 4375 or **research-ethics@griffith.edu.au**.

Feedback to you

We plan to publish the results from this study in a peer-reviewed journal and at national and international conferences. However, no personally identifiable data will be published. However If you want a copy of the results of this research, please let a member of the research team now and we will collect your contact details in order to provide you with the study results.

Expressing consent – anonymous information

Please print this sheet and retain it for your later reference. If you choose to complete the survey, you will be deemed to have consented to participate in this research.]

ONLINE CONSENT FORM

Thank you for your interest in taking part in this study, which is about developing a valuation study for a Cerebral Palsy quality of life instrument. The survey is about people preferences

when choosing between different health states. By completing this survey we assume you have consented to participate by reading the Participant Information and Consent Form.

If you decide to take part in this survey, then you will be asked to complete (i) a series of demographic questions about your child and (ii) ask what you think they would prefer when facing hypothetical scenarios. You will then be presented with a number of choice tasks, which describe two hypothetical health states.

We anticipate the full survey will take between 10-15 minutes to complete. You will be able to save and return to the survey at a later date, if required. The project is conducted by the Centre of Applied Health Economics, Griffith University, Brisbane, and is Mina Bahrampour PhD project.

It is not expected that there will be direct benefits for you participating in this survey, however, the results will be used to inform the design of health programs to support people with Cerebral Palsy. Results will be published in academic journals and conference presentations. There are no foreseeable risks associated with participation in this research. All data will be retained in a password protected electronic file at Griffith University for a period of five years before being destroyed.

So by ticking the box below and continuing to the survey, you are indicating your agreement with the following:

    i.   I have read the information about this survey.

    ii.   I understand that I am free to withdraw my participation at any point during the survey.

    iii.   I understand that participation in the survey is completely voluntary.

    iv.   I understand that responses I provide to the survey will be combined with those provided by other participants and will be statistically analysed for presentation in published reports and peer review articles, and that no personally identifiable information about me will appear in any report or article.

Please click the relevant box below to indicate your agreement and to progress to the survey:

 - I agree to participate in this study   ☐

 - I do not agree to participate in this study   ☐

If you have any questions about the research, or wish to withdraw from the study at any time, please contact Mina Bahrampour:

Email: mina.bahrampour@griffithuni.edu.au   Phone: (07) 3735 9110

Screening Questions

In this section, we would like to know about you.

1. What is your age?

2. What is your gender?

Select only one answer

Male

Female

DCE tasks:

Introducing a DCE :

In the following screens, you will be presented with 13 choice sets, which will describe hypothetical scenarios about two different hypothetical health states. You will be asked to choose the health state you prefer, based on the information provided in each choice tasks. Although each choice task may appear similar, you will see that each health state has advantages and disadvantages and you will need to carefully trade-off the advantages and disadvantages and choose the health state you prefer.

In this type of study, a person will be in the hypothetical health condition for the duration (one to ten years) as mentioned in the health condition, and then dies afterwards. Please take note of the period alive when making your decision. We know that this is a hard question to answer and we apologies for any inconvenience caused.

Please tick your preferred health state?

	Health State A	Health State B
How you feel about how you get accepted by people in general	Unhappy	Very Happy
How you feel about the way you get around	Neither Happy nor Unhappy	Very Happy
How you feel about using your hands	Very Happy	Neither Happy nor Unhappy
How you feel about the way people communicate with you	Neither Happy nor Unhappy	Very Unhappy
How you feel about the amount of pain you have	Somewhat Upset	Not at all Upset
How you feel about how you sleep	Happy	Very Happy
You will stay in this state for the following length of time, and then die	5 years	3 years
Please tick your preferred health state.	○ Health State A	○ Health State B

190

How do you feel about your **social well-being and acceptance**?

I feel very happy about how I am accepted by people in general.

I feel happy about how I am accepted by people in general.

I feel neither happy nor unhappy about how I am accepted by people in general.

I feel unhappy about how I am accepted by people in general.

I feel very unhappy about how I am accepted by people in general.

How do you feel about your **physical health**?

I feel very happy about the way I get around.

I feel happy about the way I get around.

I feel neither happy nor unhappy about the way I get around.

I feel unhappy about the way I get around.

I feel very unhappy about the way I get around.

How do you feel about your **manual ability (using your hands)**?

I feel very happy about how I use my hands.

I feel happy about how I use my hands.

I feel neither happy nor unhappy about how I use my hands.

I feel unhappy about how I use my hands.

I feel very unhappy about how I use my hands.

How do you feel about your **communication**?

I feel very happy about the way people communicate with me.

I feel happy about the way people communicate with me.

I feel neither happy nor unhappy about the way people communicate with me.

I feel unhappy about the way people communicate with me.

I feel very unhappy about the way people communicate with me.

How does you feel about your **pain and discomfort**?

I feel not at all upset about the amount of pain I have.

I feel slightly upset about the amount of pain I have.

I feel somewhat upset about the amount of pain I have.

I feel moderately upset about the amount of pain I have.

I feel very upset about the amount of pain I have.

 How do you feel about your **sleep**?

I feel very happy about how I sleep.

I feel happy about how I sleep.

I feel neither happy nor unhappy about how I sleep.

I feel unhappy about how I sleep.

I feel very unhappy about how I sleep.

Socio-demographics

Please provide information for the following questions:

1. Please provide your postcode:

2. What is your highest level of education?
Select only one answer

Year 11 or below

Year 12

Trade certificate

Diploma

Bachelor's degree

Postgraduate degree

3. Do you pay for private health insurance

Select only one answer

Yes

No

4. What best describes the total income from all adults living in your household??

Select only one answer

Less than $38,000

$38,000 to $70,000

$70,000 to $112,000

$112,000 to $205,000

More than $205,000

Prefer not to say

5. Do you have any children, If yes, please state how many?
6. If any of your children are disabled, please include the child's disability below?

AQol-4D

Tick the box that best describes your situation as it has been over the past week

aQol1. Do you need any help looking after yourself? *(For example: dressing, bathing, eating)*

☐ I need no help at all.

☐ Occasionally I need some help with personal care tasks.

☐ I need help with the more difficult personal care tasks.

☐ I need daily help with most or all personal care tasks.

aQol2. When doing household tasks: *(For example: cooking, cleaning the house, washing)*

☐ I need no help at all.

☐ Occasionally I need some help with household tasks.

☐ I need help with the more difficult household tasks.

☐ I need daily help with most or all household tasks.

aQol3. Thinking about how easily you can get around your home and community:

☐ I get around my home and community by myself without any difficulty.

☐ I find it difficult to get around my home and community by myself.

☐ I cannot get around the community by myself, but I can get around my home with some difficulty.

☐ I cannot get around either the community or my home by myself.

aQol4. Because of your health, your relationships (for example: with your friends, partner or parents) generally:

☐ Are very close and warm.

☐ Are sometimes close and warm.

☐ Are seldom close and warm.

☐ I have no close and warm relationships.

aQol5. Thinking about your relationship with other people:

☐ I have plenty of friends, and am never lonely.

☐ Although I have friends, I am occasionally lonely.

☐ I have some friends, but am often lonely for company.

☐ I am socially isolated and feel lonely.

aQol6. Thinking about your health and your relationship with your family:

☐ My role in the family is unaffected by my health.

☐ There are some parts of my family role I cannot carry out.

☐ There are many parts of my family role I cannot carry out.

☐ I cannot carry out any part of my family role.

aQol7. Thinking about your vision, including when using your glasses or contact lenses if needed:

☐ I see normally

☐ I have some difficulty focusing on things, or I do not see them sharply. *For example: small print, a newspaper or seeing objects in the distance.*

☐ I have a lot of difficulty seeing things. *My vision is blurred. For example: I can see just enough to get by with.*

☐ I only see general shapes, or am blind. *For example: I need a guide to move around.*

aQol8. Thinking about your hearing, including using your hearing aid if needed:

☐ I hear normally

☐ I have some difficulty hearing or I do not hear clearly. *For example: I ask people to speak up, or turn up the TV or radio volume.*

☐ I have difficulty hearing things clearly. *For example: Often I do not understand what is said. I usually do not take part in conversations because I cannot hear what is said.*

☐ I hear very little indeed.

For example: I cannot fully understand loud voices speaking directly to me.

aQol9. When you communicate with others: (For example: by talking, listening, writing or signing.)

☐ I have no trouble speaking to them or understanding what they are saying

☐ I have some difficulty being understood by people who do not know me. I have no trouble understanding what others are saying to me.

☐ I am only understood by people who know me well. I have great trouble understanding what others are saying to me.

☐ I cannot adequately communicate with others.

aQol10. Thinking about how you sleep:

☐ I am able to sleep without difficulty most of the time.

☐ My sleep is interrupted some of the time, but I am usually able to go back to sleep without difficulty.

☐ My sleep is interrupted most nights, but I am usually able to go back to sleep without difficulty.

☐ I sleep in short bursts only. I am awake most of the night.

aQol11. Thinking about how you generally feel:

☐ I do not feel anxious, worried or depressed.

☐ I am slightly anxious, worried or depressed.

☐ I feel moderately anxious, worried or depressed.

☐ I am extremely anxious, worried or depressed.

aQol12. How much pain or discomfort do you experience:

☐ None at all.

☐ I have moderate pain.

☐ I suffer from severe pain.

☐ I suffer unbearable pain.

On a scale of 1 to 4, how much difficulty did you have **understanding** when you were asked to choose between two scenarios (the discrete choice experiment (DCE) questions)?

1(very easy)

2 (easy)

3 (difficult)

4(very difficult)

On a scale of 1 to 4, how much difficulty did you experience when **answering** when you were asked to choose between two scenarios (the discrete choice experiment (DCE) questions)?

1(very easy)

2 (easy)

3 (difficult)

4(very difficult)

Please provide any additional comments :

CHAPTER 5:

Appendix K Utility values for the CP-6D a Cerebral palsy specific multi attribute utility instrument using Discrete Choice Experiment using mixed logit

Appendix 1: Mixed logit results

	Coefficients (standard error)	Standard deviation(standard error)
Duration	0.689(0.028) ***	0.437(0.019) ***
Acceptance × Duration		
2	-0.008(0.007)	0.093(0.014)
3	-0.089(0.008) ***	-0.026(0.022) ***
4	-0.152(0.009) ***	0.073(0.015) ***
5	-0.188(0.01) ***	0.098(0.013) ***
Physical health × Duration		
2	0.001(0.007)	-0.042(0.026)
3	-0.048(0.007) ***	0.081(0.014) ***
4	-0.134(0.008) ***	0.085(0.014) ***
5	-0.167(0.009) ***	0.068(0.016) ***
Manual ability × Duration		
2	-0.02(0.007) ***	0.051(0.017) ***
3	-0.074(0.008) ***	0.061(0.017) ***
4	-0.152(0.009) ***	0.068(0.015) ***
5	-0.197(0.01) ***	0.098(0.012) ***
Communication × Duration		
2	-0.014(0.007) **	0.006(0.035) **
3	-0.041(0.007) ***	-0.03(0.023) ***
4	-0.123(0.008) ***	0.097(0.013) ***
5	-0.144(0.008) ***	0.098(0.013) ***
Pain × Duration		
2	-0.039(0.007) ***	-0.062(0.017) ***
3	-0.051(0.007) ***	0.06(0.017) ***
4	-0.076(0.007) ***	0.006(0.035) ***
5	-0.222(0.01) ***	0.148(0.012) ***
Sleep × Duration		

2	-0.009(0.007) ***	0.074(0.015) ***
3	-0.06(0.007) ***	0.067(0.014) ***
4	-0.12(0.008) ***	0.084(0.015) ***
5	-0.171(0.01) ***	0.124(0.011) ***

Coefficients ware estimated as the interaction of that level with duration.
Levels of statistical significance (p-value): ***1%; **5%; *10%

CHAPTER 6:

Appendix L Comparing multi-attribute utility instruments: CP-6D vs AQoL-4D

Appendix 1: Ceiling and floor effect in the items:

CP-6D items	Acceptance by people	Physical health	Manual ability	Communication	Pain	Sleep
very happy	26.87	28.92	48.15	36.81	37.01	22.88
happy	44.06	39.36	35.31	40.21	32.27	30.07
neither happy nor unhappy	22.43	18.38	10.99	17.38	17.38	22.88
unhappy	5.14	10.89	4.6	4.6	8.99	18.68
very unhappy	1.5	2.45	0.95	1	4.35	5.49

AQol items	looking after yourself	Household tasks	Get around	Relation ship	Other relations	Family relation	Vision	Hearing	Comm unication	Sleep	Emotion	Pain
1	84.77	71.98	85.06	53.9	34.87	70.18	56.24	70.18	83.02	40.31	40.31	43.56
2	10.74	18.03	10.34	33.37	41.61	22.03	39.06	22.03	13.64	31.27	37.96	48.2
3	2.95	8.19	3.3	8.34	15.63	5.59	4.3	5.59	2.45	21.13	15.28	6.69
4	1.55	1.8	1.3	4.4	7.89	2.2	0.4	2.2	0.9	7.29	6.44	1.55

Apendicies References:

1.   Gamper E-M, Holzner B, King MT, Norman R, Viney R, Nerich V, et al. Test-Retest Reliability of Discrete Choice Experiment for Valuations of QLU-C10D Health States. Value in Health. 2018;21(8):958-66.
2.   Selivanova A, Buskens E, Krabbe PFM. Head-to-Head Comparison of EQ-5D-3L and EQ-5D-5L Health Values. PharmacoEconomics. 2018;36(6):715-25.
3.   Cole A, Shah K, Mulhern B, Feng Y, Devlin N. Valuing EQ-5D-5L health states 'in context' using a discrete choice experiment. European Journal of Health Economics. 2018;19(4):595-605.
4.   Craig BM, Rand K. Choice Defines QALYs: A US Valuation of the EQ-5D-5L. Med Care. 2018;56(6):529-36.
5.   Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5D-5L value set for England. Health Economics. 2018;27(1):7-22.
6.   King MT, Viney R, Simon Pickard A, Rowen D, Aaronson NK, Brazier JE, et al. Australian Utility Weights for the EORTC QLU-C10D, a Multi-Attribute Utility Instrument Derived from the Cancer-Specific Quality of Life Questionnaire, EORTC QLQ-C30. PharmacoEconomics. 2018;36(2):225-38.

7.   Rowen D, Labeit A, Stevens K, Elliott J, Mulhern B, Carlton J, et al. Estimating a Preference-Based Single Index Measuring the Quality-of-Life Impact of Self-Management for Diabetes. Medical Decision Making. 2018;38(6):699-707.

8.   Jonker MF, Donkers B, de Bekker-Grob EW, Stolk EA. Advocating a Paradigm Shift in Health-State Valuations: The Estimation of Time-Preference Corrected QALY Tariffs. Value in Health. 2018;21(8):993-1001.

9.   Krabbe PFM, Stolk EA, Devlin NJ, Xie F, Quik EH, Pickard AS. Head-to-head comparison of health-state values derived by a probabilistic choice model and scores on a visual analogue scale. European Journal of Health Economics. 2017;18(8):967-77.

10.   Xie F, Pullenayegum E, Pickard AS, Ramos Goñi JM, Jo Mw, Igarashi A, et al. Transforming Latent Utilities to Health Utilities: East Does Not Meet West. Health Economics. 2017;26(12):1524-33.

11.   Purba F, Hunfeld J, Iskandarsyah A, Fitriana T, Sadarjoen S, Ramos-Goñi J, et al. The Indonesian EQ-5D-5L Value Set. PharmacoEconomics. 2017;35(11):1153-65.

12.   Huynh E, Coast J, Rose J, Kinghorn P, Flynn T. Values for the ICECAP-Supportive Care Measure (ICECAP-SCM) for use in economic evaluation at end of life. Social Science & Medicine. 2017;189:114-28.

13.   Mulhern B, Bansback N, Hole AR, Tsuchiya A. Using Discrete Choice Experiments with Duration to Model EQ-5D-5L Health State Preferences. Medical Decision Making. 2017;37(3):285-97.

14.   Ludwig K, von der Schulenburg JG, Greiner W. Valuation of the EQ-5D-5L with composite time trade-off for the German population - an exploratory study. Health & Quality of Life Outcomes. 2017;15(1):39.

15.   Ramos-Goñi JM, Pinto-Prades JL, Oppe M, Cabasés JM, Serrano-Aguilar P, Rivero-Arias O. Valuation and modeling of EQ-5D-5L health states using a hybrid approach. Med Care. 2017;55(7):e51-e8.

16.   Mulhern B, Norman R, Lorgelly P, Lancsar E, Ratcliffe J, Brazier J, et al. Is Dimension Order Important when Valuing Health States Using Discrete Choice Experiments Including Duration? Pharmacoeconomics [Internet]. 2016:[1-13 pp.]. Available from: https://link.springer.com/article/10.1007%2Fs40273-016-0475-z.

17.   Mulhern B, Shah K, Janssen MF, Longworth L, Ibbotson R. Valuing Health Using Time Trade-Off and Discrete Choice Experiment Methods: Does Dimension Order Impact on Health State Values? Value in Health. 2016;19(2):210-7.

18.   Norman R, Kemmler G, Viney R, Pickard AS, Gamper E, Holzner B, et al. Order of Presentation of Dimensions Does Not Systematically Bias Utility Weights from a Discrete Choice Experiment. Value in Health. 2016;19(8):1033-8.

19.   Norman R, Viney R, Aaronson N, Brazier J, Cella D, Costa D, et al. Using a discrete choice experiment to value the QLU-C10D: feasibility and sensitivity to presentation format. Quality of Life Research. 2016;25(3):637-49.

20.   Rowen D, Brazier J, Mukuria C, Keetharuth A, Risa Hole A, Tsuchiya A, et al. Eliciting Societal Preferences for Weighting QALYs for Burden of Illness and End of Life. Medical Decision Making. 2016;36(2):210-22.

21.   Shiroiwa T, Ikeda S, Noto S, Igarashi A, Fukuda T, Saito S, et al. Comparison of Value Set Based on DCE and/or TTO Data: Scoring for EQ-5D-5L Health States in Japan. Value in Health. 2016;19(5):648-54.

22.   Scalone L, Stalmeier PF, Milani S, Krabbe PF. Values for health states with different life durations. European Journal of Health Economics. 2015;16(9):917-25.

23.   Eckert J, Lichters M, Piedmont S, Vogt B, Robra BP. Case vignettes based on EQ-5D to elicit stated preferences for health services utilization from the insurees' perspective. BMC Health Serv Res. 2015;15:481.

24.   Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. Medical Decision Making. 2015;35(3):328-40.

25.   Bansback N, Hole AR, Mulhern B, Tsuchiya A. Testing a discrete choice experiment including duration to value health states for large descriptive systems: addressing design and sampling issues. Social Science & Medicine. 2014;114:38-48.

26.   Krabbe PF, Devlin NJ, Stolk EA, Shah KK, Oppe M, van Hout B, et al. Multinational evidence of the applicability and robustness of discrete choice modeling for deriving EQ-5D-5L health-state values. Med Care. 2014;52(11):935-43.

27.   Norman R, Viney R, Brazier J, Burgess L, Cronin P, King M, et al. Valuing SF-6D Health States Using a Discrete Choice Experiment. Medical Decision Making. 2014;34(6):773-86.

28.   Viney R, Norman R, Brazier J, Cronin P, King MT, Ratcliffe J, et al. An Australian discrete choice experiment to value eq-5d health states. Health Economics. 2014;23(6):729-42.

29.   Norman R, Cronin P, Viney R. A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states. Appl Health Econ Health Policy. 2013;11(3):287-98.

30.   Ramos-Goni JM, Rivero-Arias O, Errea M, Stolk EA, Herdman M, Cabases JM. Dealing with the health state 'dead' when using discrete choice experiments to obtain values for EQ-5D-5L heath states. European Journal of Health Economics. 2013;14 Suppl 1:S33-42.

31.   Prosser LA, Payne K, Rusinak D, Shi P, Messonnier M. Using a discrete choice experiment to elicit time trade-off and willingness-to-pay amounts for influenza health-related quality of life at different ages. Pharmacoeconomics. 2013;31(4):305-15.

32.   Pullenayegum E, Xie F. Scoring the 5-level EQ-5D: can latent utilities derived from a discrete choice model be transformed to health utilities derived from time tradeoff tasks? Medical Decision Making. 2013;33(4):567-78.

33.   Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. Journal of Health Economics. 2012;31(1):306-18.

34.   Ratcliffe J, Flynn T, Terlich F, Stevens K, Brazier J, Sawyer M. Developing adolescent-specific health state values for economic evaluation: an application of profile case best-worst scaling to the Child Health Utility 9D. Pharmacoeconomics. 2012;30(8):713-27.

35.   Brazier J, Rowen D, Yang Y, Tsuchiya A. Comparison of health state utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale. European Journal of Health Economics. 2012;13(5):575-87.

36.   Stolk EA, Oppe M, Scalone L, Krabbe PF. Discrete choice modeling for the quantification of health states: the case of the EQ-5D. Value in Health. 2010;13(8):1005-13.

37.   Ratcliffe J, Brazier J, Tsuchiya A, Symonds T, Brown M. Using DCE and ranking data to estimate cardinal values for health states for deriving a preference-based single index from the sexual quality of life questionnaire. Health economics. 2009;18(11):1261-76.

38.   Burr JM, Kilonzo M, Vale L, Ryan M. Developing a preference-based Glaucoma Utility Index using a discrete choice experiment. Optom Vis Sci. 2007;84(8):797-808.

39.   Bryant FB, Yarnold PR. Principal-components analysis and exploratory and confirmatory factor analysis. 1995.

40.   Jolliffe IT. Principal component analysis and factor analysis. Principal component analysis: Springer; 1986. p. 115-28.

41.   Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis and rheumatism. 2007;57(8):1358-62.

42.   Jabrayilov R, Emons WH, Sijtsma K. Comparison of classical test theory and item response theory in individual change assessment. Applied Psychological Measurement. 2016;40(8):559-72.