VALUATION OF EQ-5D-5L STUDY FOR THE MALAYSIAN POPULATION

ANNUSHIAH A/P VASAN THAKUMAR

UNIVERSITI SAINS MALAYSIA

2020

VALUATION OF EQ-5D-5L STUDY FOR THE MALAYSIAN POPULATION

by

ANNUSHIAH A/P VASAN THAKUMAR

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

July 2020

ACKNOWLEDGEMENT

Firstly, I would like to express gratitude to my main supervisor, Associate Professor Dr Asrul Akmal Shafie for believing in me, providing so much support during this five-year journey, and granting me the freedom to express my creativity and ideas for the project to blossom. My two co-supervisors, Dr Lim Ching Jou and Associate Professor Dr Luo Nan were ever so helpful in lending a hand with research advice, supporting the research with their respective expertise, and easing my PhD journey with their kindness and motivation. To Dr Asrul, Dr Lim, and Dr Nan, your work ethics, commitment, and enthusiasm are exemplary, and I hope to walk in your shoes one day.

Next, this journey would not have been possible without funding from Universiti Sains Malaysia's Research University grants (1001/PFARMASI/811287), (1001/PFARMASI/816114), and EuroQol Research Foundation (EQ Project 2016100). The ideas for an impactful research materialized in-part because these organizations believed in the study and for that thank you very much! Additionally, the funding received allowed 18 interviewers to be hired for the study, whom I am blessed to have worked and travel with around Malaysia for the data collection phase. They taught me the importance and impact of teamwork and dedication.

Also, not forgetting the staff and students at the Discipline of Social and Administrative Pharmacy, USM who were kind and helpful. They made learning fun! The four friendships fostered at DSAP has made this journey especially meaningful and these people remind me that one can excel in life while being kind and empathetic at the same time. These dear friends are Siti Fauziah Abu, Jacqueline Wong Hui Yi, Noor Syahireen Mohammed, and Irwinder Kaur Chhabra. Not forgetting Haarathi

Chandriah, Sivaraj Raman, Khairu Hazwan Mustaffa, Tuqa Haitham, Ho Rhu Yann, Mohammed Firdauz Isahak, and the rest of the DSAP gang too for being so supportive and helpful during this journey. I am thankful to have meet these wonderful people.

I would also like to extend my gratitude to the examiners of my PhD viva voce for taking time and effort to thoroughly read my thesis and prove useful suggestions in better capturing the study undertaken. The reviewers of my articles and presentations have also provided invaluable suggestions for improvement.

Last but not least, this journey materialized with the blessings and emotional support from my mother, father, my brother, and my two cats, Catherine and Simba. I am who I am today mainly because of their faith in me. They motivate me to be a better person every day. I am blessed to have so many kind beings in my life who have eased my PhD journey and from whom I have learnt so much, and for that I thank God.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AD	anxiety dimension
ADD10_TTO	additive 10-parameter time trade-off model
ADD10_VAS	additive 10-parameter visual analogue scale model
ADD11_VAS	additive 11-parameter visual analogue scale model
ADD20_HYBRID	additive 20-parameter hybrid model
ADD20_TTO	additive 20-parameter time trade-off model
AIC	Akaike information criterion
AQoL	Assessment of Quality of Life
BIC	Bayesian information criterion
BTD	better than dead
CAPI	computer-administered personal interviews
CCC	concordance correlation coefficient
COSMIN	Consensus-based standards for the selection of health measurement instruments
C-TTO	composite time trade-off
CUA	cost-utility analysis
D	dead
DCE	discrete choice experiment
EQ-5D	EuroQol 5-dimensional questionnaire
EQ-5D-3L/3L	Three-level EuroQol five-dimensional questionnaire
EQ-5D-5L/5L	Five-level EuroQol five-dimensional questionnaire
EQ-VAS	EuroQol visual analogue scale
EQ-VT	EuroQol Valuation Technology
EUT	Expected Utility Theory
FGDs	focus group discussions
GLS	generalized least squares
H’	Shannon entropy/ Shannon index
HRQoL	health-related quality of life
HTA	health technology assessments
HUI	Health Utilities Index
ICC	intra-class correlation coefficient
J’	Shannon evenness index

MAE	mean absolute error
MaHTAS	Malaysian Health Technology Assessment Section
MO	mobility dimension
MOH	Ministry of Health Malaysia
MOHMF	Ministry of Health Malaysia Medicines Formulary
MSE	mean square error
MULT6_TTO	multiplicative 6-parameter time trade-off model
MULT6_VAS	multiplicative 6-parameter visual analogue scale model
MULT7_TTO	multiplicative 7-parameter time trade-off model
MULT8_HYBRID	multiplicative 8-parameter hybrid model
MULT8_TTO	multiplicative 8-parameter time trade-off model
MVH	Measurement and Valuation of Health valuation study
NICE	National Institute of Health and Care
OLS	ordinal least squares
PBMs	preference-based measures
PD	pain/discomfort dimension
PPDD	Formulary Management Branch of Pharmacy Practice & Development Division
PRR	private engagement in religiosity
PUR	public engagement in religiosity
QALYs	quality-adjusted life years
QC	quality control
QWB	Quality of Well-Being Scale
R^2	Pearson's R
RUM	random utility maximization
SC	self-care dimension
SD	standard deviation
SE	standard error
SF-36D	36-Item Short Form Survey
SF-6D	Six-Item Short Form Survey
SG	standard gamble
SRR	self-reported religiosity
TTO	time trade-off
UA	usual activities dimension
VAS	visual analogue scale

WHO	World Health Organization
WTD	worse than dead

LIST OF APPENDICES

KAJIAN PENILAIAN EQ-5D-5L BAGI MASYARAKAT MALAYSIA

ABSTRAK

Peningkatan kos perbelanjaan penjagaan kesihatan mendorong perancangan peruntukan sumber yang berhemat serta mengukuhkan peranan penilaian ekonomi dalam membimbing keputusan sebegini. Analisis kos-utiliti adalah salah satu alternatif dimana hasil diukur menggunakan jumlah tahun yang disesuaikan dengan kualiti (QALYs). Keutamaan kesihatan difaktorkan ke dalam pengiraan QALYs dan instrumen berasaskan keutamaan generik (PBMs) biasa digunakan untuk menjana pemberat ini. Antara PBM yang paling kerap digunakan untuk memaklumkan keputusan penilaian ekonomi di seluruh dunia ialah EQ-5D-5L. Walau bagaimanapun, kajian pengesahan EQ-5D-5L serta penjanaan set nilai khusus kepada negara Malaysia tidak pernah dijalankan. Oleh itu, objektif utama tesis ini adalah untuk menganggarkan suatu set nilai EQ-5D-5L yang sah untuk masyarakat Malaysia. Dua perbincangan kumpulan berfokus (FGD) telah diadakan untuk meneroka pemahaman masyarakat Malaysia terhadap dimensi-dimensi EQ-5D-5L serta menyiasat dimensi tambahan (*bolt-on*) dalam melengkapi konsep kesihatan. Selanjutnya, dengan menggunakan 100 responden persampelan mudah, pengaruh dimensi tambahan pada dimensi EQ-5D-5L diuji dalam kajian perintis di Pulau Pinang. Selepas itu, menggunakan reka bentuk pensampelan pelbagai peringkat, kajian keratan rentas penilaian EQ-5D-5L pada peringkat kebangsaan dijalankan oleh 18 penemuduga terlatih ke atas 1137 populasi umum Malaysia. Ciri-ciri psikometrik EQ-5D-5L juga dinilai dalam kajian tersebut. Pelbagai kaedah pemodelan diuji pada data penilaian dan pelbagai teknik mapan digunakan untuk mengkaji prestasi model-model ini dalam meramalkan nilai-nilai keadaan kesihatan. Satu lagi set nilai dianggarkan, tetapi menggunakan versi instrumen terdahulu, EQ-5D-3L. Data ini dikumpulkan di kawasan utara Malaysia menggunakan reka bentuk pensampelan yang serupa, dengan menemubual 638 responden. Set nilai EQ-5D-5L dan EQ-5D-3L kemudiannya dibandingkan dari segi ciri-ciri model. Akhir sekali, impak faktor-faktor sosiodemografi dan keagamaan ke atas corak-corak nilai utiliti daripada set nilai anggaran EQ-5D-5L dinilai dengan menggunakan kaedah regresi. Hasil FGD menunjukkan bahawa masyarakat Malaysia umumnya memahami dimensi-dimensi EQ-5D-5L dengan baik dan 11 bolt-on telah dikenalpasti dan menunjukkan potensi untuk kajian selanjutnya. Versi EQ-5D-5L Bahasa Melayu dan Inggeris memaparkan pengurangan kesan siling, kesahihan konvergen yang boleh diterima, dan keputusan kebolehpercayaan ujian-uji semula yang lemah sehingga sederhana. Hasil pemodelan data penilaian menunjukkan bahawa model terbaik ialah model berdaya darab hibrid 8-parameter dengan set nilai antara -0.442 hingga 1. Perbandingan dengan model EQ-5D-3L memaparkan nilai-nilai EQ-5D-5L yang lebih luas dan analisis ke atas pola utiliti mendedahkan bahawa faktor-faktor sosiodemografi mempunyai impak ke atas set nilai EQ-5D-5L. Kesimpulannya, set nilai EQ-5D-5L yang khusus untuk Malaysia telah dianggarkan menggunakan metodologi yang mantap.

VALUATION OF EQ-5D-5L STUDY FOR THE MALAYSIAN POPULATION

ABSTRACT

With the rise of healthcare expenditure, prudent resource allocation decisions become vital and economic evaluations offer valuable information in guiding these decisions. Cost-utility analysis is one such alternative with outcomes measured using quality-adjusted life years (QALYs). QALYs takes into consideration health preference weights, with generic preference-based measures (PBMs) being most commonly used to generate these weights. One of the most widely used instrument to-date, EQ-5D-5L has informed economic evaluation decisions throughout the world. However, Malaysia has yet to validate the measure and apply it to generate a country-specific value set. Therefore, the main objective of this thesis is to estimate a valid EQ-5D-5L value set for Malaysians. Two focus group discussions (FGDs) were held to explore the EQ-5D-5L dimensions understanding in Malaysians and investigate potential dimensions (bolt-ons) that may complement the concept of health. Next, using 100 conveniently sampled respondents, the influence of these identified bolt-ons on the current EQ-5D-5L dimensions were tested in a pilot setting in Penang. Subsequently, employing a cross-sectional multi-stage sampling design, an EQ-5D-5L nationwide valuation study was carried out by 18 trained interviewers on 1137 Malaysian general population. The psychometric properties of EQ-5D-5L were also assessed in the process. A variety of modelling methods were fitted onto the valuation data and performance of these models in predicting health state values were assessed using robust techniques. Another value set was estimated, but applying an earlier version of the instrument, EQ-5D-3L. Data was collected in the Northern regions of

Malaysia using a similar sampling design employing 638 respondents. The EQ-5D-5L and EQ-5D-3L value sets were then compared in terms of model characteristics. Lastly, using utility values from EQ-5D-5L estimated value set, the impact of sociodemographic variables and religiosity were on these patterns were assessed using regression methods. FGD results showed that Malaysians generally perceived the EQ-5D-5L dimensions well and 11 bolt-ons were identified and pilot tested with some showing potential for further testing. The Malay and English EQ-5D-5L versions displayed reduced ceiling effect, acceptable convergent validity, but poor to moderate test-retest reliability results. Valuation data modelling revealed the best performing model was the multiplicative 8-parameter hybrid model with a value set range of -0.442 to 1. Comparisons with the EQ-5D-3L models demonstrated the wider range of values that EQ-5D-5L possesses, with analysis into the utility patterns revealing sociodemographic factors did impact the results. In conclusion, a Malaysian-specific valid EQ-5D-5L value set was estimated using a robust methodology.

CHAPTER 1

INTRODUCTION

1.1 Cost-utility analysis and QALYs

Economic evaluations are increasingly applied in guiding health care resource allocation decisions around the world. Cost-utility analysis (CUA) is a type of economic evaluation that captures health benefits or outcomes in the form of quality-adjusted life years (QALYs) (Drummond, Sculpher, Claxton, Stoddart, & Torrance, 2015b). Incorporating both the elements of quality (health preference measured as utility values) and quantity (amount of life years), QALYs offer a wholesome approach to informing economic evaluations. As utility values can be calculated for a variety of morbidities, QALYs are not limited to comparators within a disease.

QALYs is the most frequently recommended form of outcomes measure in pharmacoeconomic guidelines around the world, and includes countries such as Ireland, Canada, New Zealand, Taiwan, Japan, and Thailand (Chaikledkaew & Kittrongsiri, 2014; Eldessouki & Smith, 2012; Griffiths, Legood, & Pitt, 2016). Additionally, health technology bodies such as the National Institute for Health and Care Excellence (NICE) and the Second Panel on Cost-Effectiveness in Health and Medicine advocate the use of QALYs in valuing health benefits for technology assessments including health technologies and medical procedures (Earnshaw & Lewis, 2008; National Institute for Health and Care Excellence, 2013; Sanders et al., 2016).

1.2 Health preference

Health-related quality of life (HRQoL) measures are complementary to the conventional measure of clinical indicators for an in depth of a person's well-being

(Fayers & Machin, 2013). Preference-based measures (PBMs) are a category of HRQoL instruments that additionally evaluates the health preference of an individual.

Health preference data is essentially applied in estimating value sets and consequently used in calculating the 'quality' component of QALYs (Brazier, Ratcliffe, Salomon, & Tsuchiya, 2007b). While there are many ways to generate utility values, one of the preferred alternatives is to use a generic PBM such as the EuroQol five-dimensional questionnaire (EQ-5D). Estimating value sets (comprising utility values for all health states that the instrument can define) from generic PBMs facilitate standardization and increase comparability between studies.

As CUA continually expands as an economic evaluation tool, the use of PBMs and the availability of value sets become almost a necessity. Cross-country differences in terms of health preference have been shown to exist and these differences subsequently lead to contrasting dimensions and diseases being prioritized for healthcare improvements. Thus, countries should preferably apply their own value sets to reflect the health preference patterns of their respective populations (Mahlich, Dilokthornsakul, Sruamsiri, & Chaiyakunapruk, 2018; Ombler, Albert, & Hansen, 2018; Wang et al., 2019).

1.3 Measuring health preference

According to the World Health Organization (WHO), health is defined as 'A state of complete physical, mental, and social well-being, and not merely the absence of disease and infirmity' (World Health Organization, 1948). Concurring with WHO's definition of health, an individual's experience of a health state may be subjective and more multi-dimensional than from a healthcare professional's perspective (Celli et al., 2017). It might also include elements that are difficult to measure within the scope of biomedical measurements, such feelings of pain and the ability to function in society. The availability of self-reported measures to assess health grants the individual the capacity to describe health from his or her actual perspective and experience thus providing a more comprehensive picture of one's well-being (Fitzpatrick, Davey, Buxton, & Jones, 1998; Holmes, Stanescu, & Bishop, 2019).

Recognizing the importance of the individual's health perspective, the measurement of health preferences has been embedded in healthcare decision making. In this context, health preferences of individuals are expressed in QALYs and is the outcomes measure of CUA (Drummond, Sculpher, Claxton, Stoddart, & Torrance, 2015a). Incorporating the health preference of individuals for QALY calculations requires that health be valued on a scale with 0 and 1 representing death and full health respectively (Weinstein, Torrance, & McGuire, 2009). Negative values can be present and are considered worse than dead.

These health state values can be elicited by members of the general public (using hypothetical health states) or patients themselves (using their own experiences) (Brazier et al., 2018; Rand-Hendriksen, Augestad, Kristiansen, & Stavem, 2012). The general population basically have no vested interest in particular health states and are less likely to influence the outcomes of values to their favour (such as ensuring access to a particular treatment). Another argument include that the general public are taxpayers and public preference should rightfully inform resource allocation decisions (Fryback, 2003; Menzel, Dolan, Richardson, & Olsen, 2002; Ubel, Loewenstein, & Jepson, 2003). By contrast, health states may be poorly described and the respondents who have little to no experience and might have a difficult time imagining such health conditions (Brazier et al., 2018; Michel, Engel, Rand-Hendriksen, Augestad, & Whitehurst, 2016).

While it is the well-being of patients we are most concerned with during assessments of a new treatment or health technology, patients exhibiting adaptations to their conditions may result in higher health state values being assigned to an impaired health state (Dolan, 2011; McTaggart-Cowan, Tsuchiya, O'Cathain, & Brazier, 2011) . Consequently, the incremental gain from treatments to HRQoL would possibly be of a lesser value than when the general public's preferences are elicited, thus implicating smaller improvements to the treatment and making them less desirable (Brazier, Ratcliffe, Salomon, & Tsuchiya, 2007c).

A study applying the experience-based values of 115,206 knee replacement patients found that these values are not stable and are influenced by the timing of when the patient's health was valued (Pickard, Hung, Lin, & Lee, 2017). The values of patients were assessed before and after knee replacement surgery using EQ-5D-3L. Using the 90,450 complete data available, these were modelled to assess the preference weights of each of the five dimensions pre- and post-surgery. The study found that not only that these weights increased post-surgery, the dimensional ranking of the five dimensions also altered. In other words, the same health states carried different values even when they were each rating their own health. The susceptibility of experienced-based values to the context in which the values were elicited makes it less desirable than to value hypothetical health states, which is commonly carried out using the general population. Additionally, valuation tasks used to elicit health preference are complex and may be intrusive, and patients, who are already unwell, may not be willing or able to answer them, with ethical issues raised too (Brazier et al., 2007c).

In summary, while the general public may sometimes underestimate these values, the argument that they have no vested interest in the outcome of such studies and are tax-payers funding the public healthcare are favoured over the direct valuations from patients who may adapted to the disease or exhibit unstable values. Thus, general population values are commonly use to generate country-specific value sets and are specified by agencies such as NICE in the UK for informing HTA (Brazier & Longworth, 2011; Xie, Gaebel, Perampaladas, Doble, & Pullenayegum, 2014). Such standardizations also facilitate comparability of values between different studies.

1.4 Valuation techniques

There are a number of preference-based valuation techniques that have been refined for use in healthcare research. Four commonly used ones are the visual analogue scale, standard gamble, time trade-off, and discrete choice experiments (Brazier & Ratcliffe, 2016).

1.4.1 Visual analogue scale (VAS)

The visual analogue scale (VAS) is essentially a line on which the respondents rate the preference or value of a health state, as outlined in Figure 1.1 (Krabbe, Stalmeier, Lamers, & Busschbach, 2006). Distinct end-points such as ‘best imaginable health’ and ‘worst imaginable health’ are used and the distances between the health state ratings should reflect the relative differences in the concept measured (Gudex, Dolan, Kind, & Williams, 1996). VAS is generally regarded as the simplest to complete among the valuation techniques available.

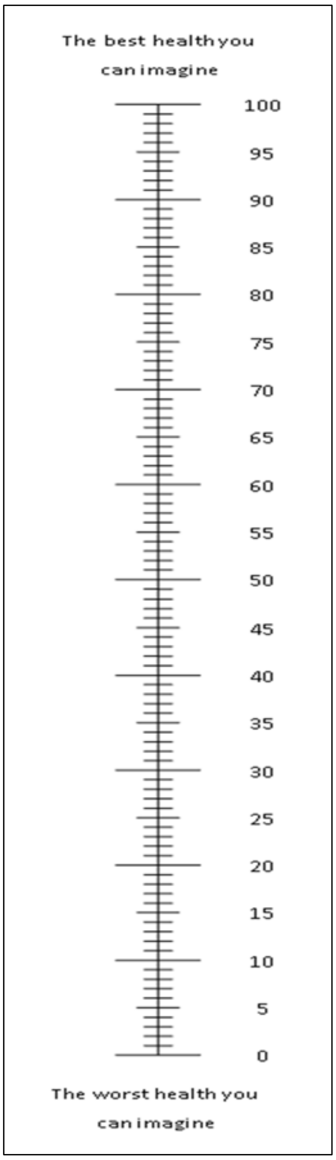


Figure 1.1 An example of a visual analogue scale, with zero as the lower endpoint and 100 as the upper endpoint

1.4.2 Standard gamble (SG)

The standard gamble (SG) technique is a choice-based task that asks the respondent to select the better between two hypothetical lives (Figure 1.2). One alternative involves an intermediate outcome (the health state to be valued, STATE *i*) that occurs with certainty and the other alternative is a gamble between a better (HEALTHY) and worse outcome (STATE *j*) with varying probabilities. The probabilities of the gamble (between the better and worse outcome) are varied until an indifference between the two alternatives (the certain and the gamble) are reached (Gudex, 1994a).

SG has its theoretical basis in Expected Utility Theory (EUT). Briefly, EUT postulates that individuals select between alternatives that maximize their expected utility (von Neumann & Morgenstern, 1944). Utility values are estimated for each possible outcome of a given alternative and multiplied to the probability of these outcomes occurring, resulting in an expected utility for that particular alternative. This process is repeated for all the alternatives available.

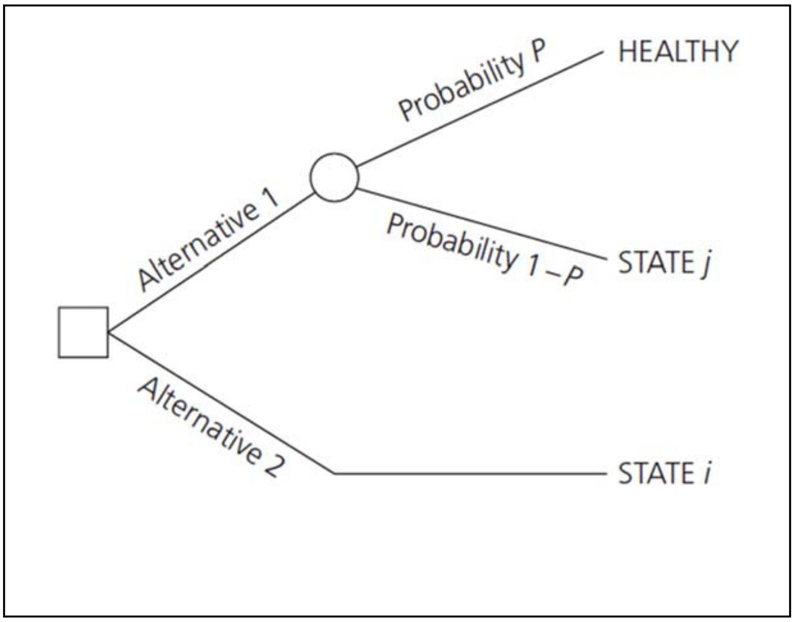


Figure 1.2 An illustration of the standard gamble technique

1.4.3 Time trade-off (TTO)

Developed as an alternative to SG, the time trade-off technique is also a choice-based task that involves selecting between two alternatives, but instead of probabilities, length of life is traded (Torrance, Thomas, & Sackett, 1972). Values are anchored on full health and death. The respondent chooses between living a longer life in a health state and living a shorter life in better health (Figure 1.3). The amount of time in better health is varied until preferential indifference between the two health states is achieved.

1.4.4 Discrete choice experiment (DCE)

Another valuation technique that applies choice-based tasks is the discrete choice experiment technique (DCE). Besides valuing health benefits, DCE is commonly applied to value ‘non-health benefits’ in healthcare (Mulhern et al., 2019). The most basic form of DCE requires respondents to state which between the two health states is preferred without probabilities or duration. DCE is a relatively newer method compared to the rest of the valuation techniques, but are increasingly used for preference elicitation in healthcare (Clark, Determann, Petrou, Moro, & de Bekker-Grob, 2014).

The concept of DCE originates from Lancaster’s economic theory of value that assumes the underlying attributes or dimensions of a health state determines the individual’s value of the health state and that health preferences are elicited through the respondents’ choices (Lancaster, 1966). Preference is then modelled under the random utility maximization (RUM) framework. Introduced in the field of

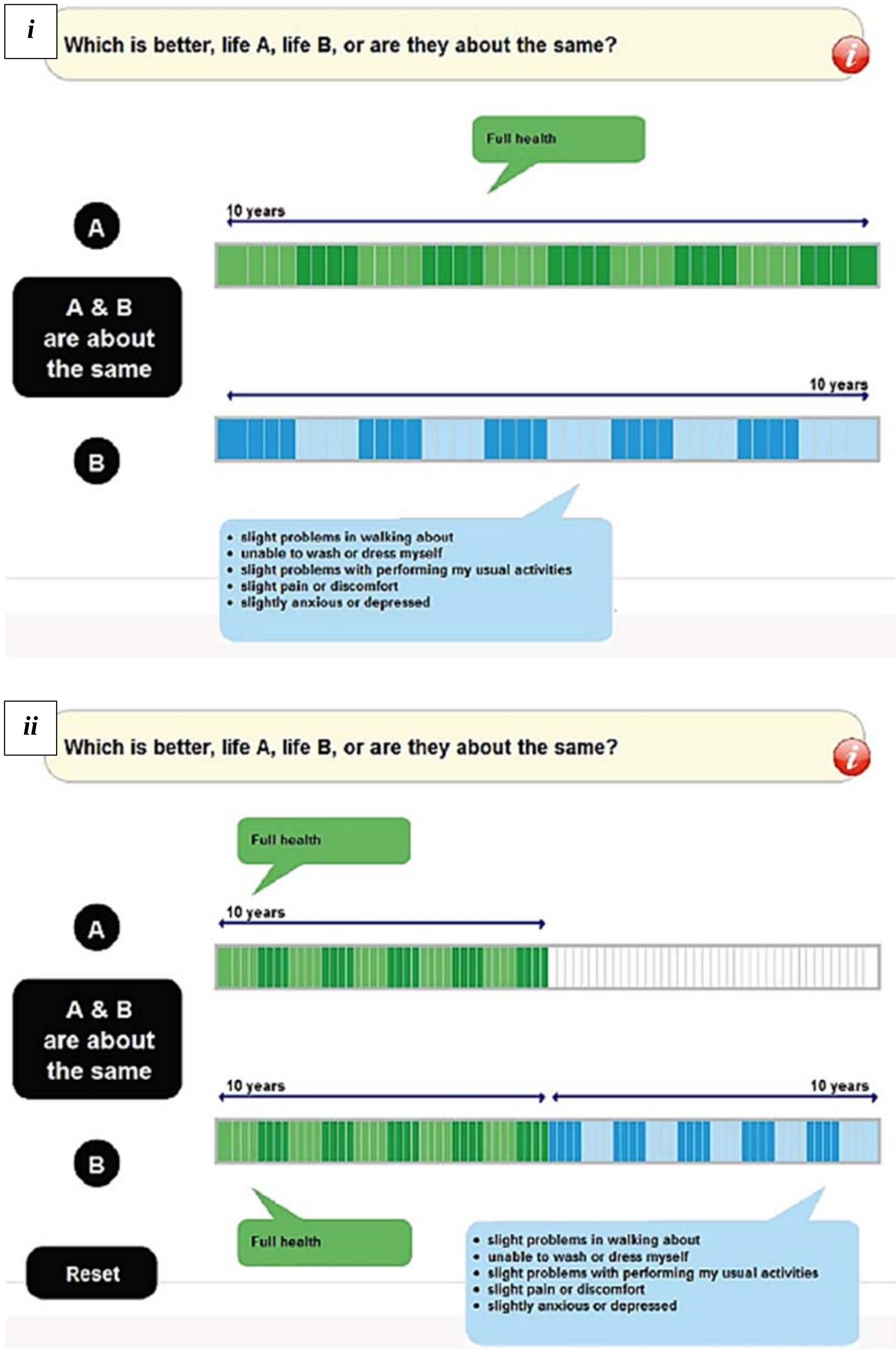


Figure 1.3 Time trade-off task illustrated using better than dead example

Note: ***i*** indicates task for better than dead states and ***ii*** for states worse than dead

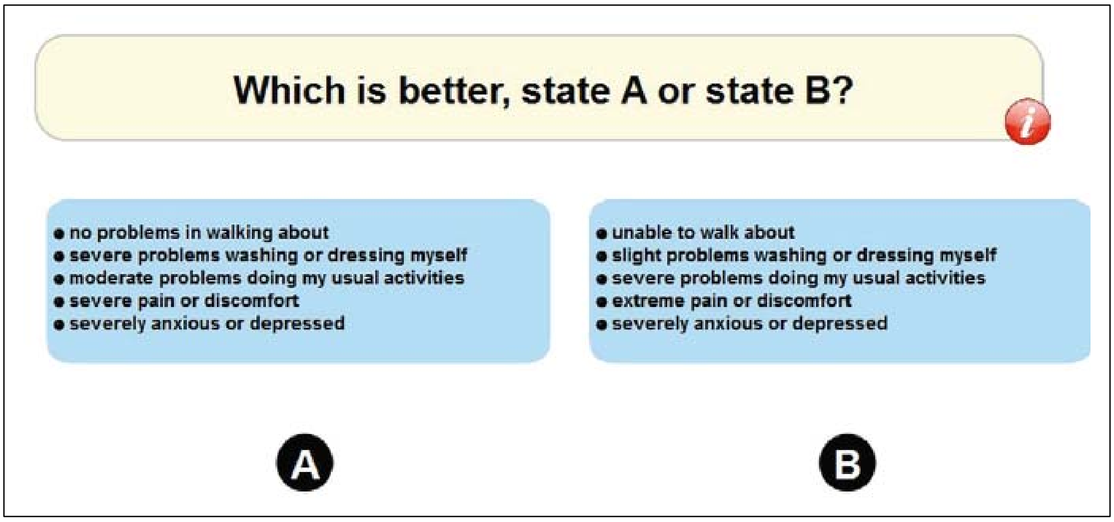


Figure 1.4 Discrete choice experiment task example

psychology, it was then applied in economics, with further developments applied later (McFadden, 1974; Thurstone, 1927).

In RUM, individuals are assumed to maximize their indirect utility through the choices made but also postulates that choice behaviour is intrinsically probabilistic (random). An individual's utility value is assumed to have two parts, an explainable, systematic utility component and a random utility component. This random component represents the unmeasured variation in preference due to unobserved factors affecting choice, differences among individual likings, or measurement errors. The probability that individuals chooses health state A over B equals the probability that the systematic utility difference between A minus B is higher than the difference between the random utility of the difference between A minus B. (Amaya-Amaya, Ryan, & Gerard, 2008).

1.4.5 Comparison of techniques

These valuation techniques have demonstrated empirical evidence in capturing the health preference of individuals (Brazier, Deverill, & Green, 1999; Craig, Busschbach, & Salomon, 2009; Green, Brazier, & Deverill, 2000; Krabbe, 2008; Pullenayegum &

Xie, 2013; Stolk, Oppe, Scalone, & Krabbe, 2010). However, they may vary in terms of their foundations, administrative burden, and concerns raised on the elicitation methods.

1.4.5(a) Theoretical leanings of valuation techniques

Some valuation techniques are more strongly rooted in theoretical foundations than others. For example, VAS is not choice-based and not rooted in any of the utility theories used in informing healthcare decision making (Bleichrodt & Johannesson, 1997).

On the other hand, SG has its basis in EUT because health preference is elicited under the condition of uncertainty (Gafni, 1994). Although it is considered as the ’gold standard’ by some due to the similar uncertain nature in medical decision making, in practice, there are evidence of EUT violations of SG (Drummond et al., 2015b; Green et al., 2000; Hershey, Kunreuther, & Schoemaker, 1982; Schoemaker, 1982).

Although decisions are made under certainty, TTO is choice-based and the methods slightly resemble that of SG. However, methods to adjust for uncertainty are present but seldom used (Stiggelbout et al., 1994). Nonetheless, TTO and SG are the recommended preference elicitation methods of many HTA agencies (Rowen, Azzabi Zouraq, Chevrou-Severac, & van Hout, 2017).

DCE, while not as extensively used to inform valuation studies as VAS, SG, and TTO, has a strong basis in RUM (Louviere & Lancsar, 2009). Additionally, the utility derived are related only to the attractiveness of the health state and not masked by underlying risk or time preference patterns commonly associated with SG and TTO tasks (Stolk et al., 2010).

1.4.5(b) Burden of administration

VAS is the simplest of valuation techniques available and is cognitively not burdensome. However, emerging research in Asia reveal that Asians face greater difficulty in rating VAS than Westerners which was associated with the lower levels of education (Cheung & Thumboo, 2006; Qian, Tan, Chuang, & Luo, 2019). In a separate qualitative study conducted in Singapore, the upper endpoint on the VAS scale labelled as 'best imaginable health' had varied interpretations of the term, leading to questions about the comparability of these values (Tan, G.L., M., & Luo, 2019). Thus, cross-country differences in education levels and the vagueness of the endpoints as interpreted by Asian communities may contribute to different comprehension levels of the VAS-based valuation technique, subsequently producing values which are not aligned with actual health preference patterns of the population.

Generally, SG-based studies has been shown to exhibit practicality in terms of high completion rates and is acceptable in different disease areas when applied in conducting disease-specific valuations for conditions in which generic PBMs may not exhibit strong psychometric validity. (Brazier et al., 2007c). The concept of gamble or probabilities is complex and may be difficult to grasp, leading to the development of TTO for healthcare (Torrance, 1976).

The acceptability and practicality of TTO has been demonstrated in the past (Green et al., 2000). However, it is also cognitively challenging to understand the concept of trade-off in TTO and proper interviewer training is essential for effective valuations to take place (Oppe, Rand-Hendriksen, Shah, Ramos‐Goñi, & Luo, 2016). Furthermore, the concept of trading life and dealing with death may become an issue, especially in Asian countries where it is associated with taboo to bring up such a topic

(Purba, Hunfeld, Iskandarsyah, Fitriana, Sadarjoen, Passchier, et al., 2017; Wee et al., 2006).

As for DCE, making a choice between two health states without having to consider probabilities and trading off time is cognitively less challenging than SG and TTO. The simplicity of the task administration allows the evaluation using a self-completion format. Indeed, an EQ-5D-5L valuation study was conducted in US involving 8222 respondents administered online DCE tasks resulting in logically consistent values (Craig & Rand, 2018).

1.4.5(c) Concerns with value elicitation

Prone to context effects, studies have shown the values of VAS health states to be affected by values of the other states on the scale and health state on the top and bottom of the scale are placed further than when direct comparisons are made, leading to an end-point bias (Robinson, Loomes, & Jones-Lee, 2001; Torrance, Feeny, & Furlong, 2001).

By contrast, there have been arguments that other factors such as loss aversion and attitude to risk influence SG valuations, and thus may not truly represent actual health preference (Broome, 1993; Lipman, Brouwer, & Attema, 2019; Richardson, 1994). In other words, people may be so averse to loss of good health that they constantly prefer to live in the 'healthy' alternative rather than the health state to be valued, even if the probability in the 'healthy' state is very low, leading to the health state in question being valued lower than necessary. In contrast, those who have a low attitude? to risk will value a health state higher than those who are risk-takers as they are not willing to gamble for a better health state.

On the other hand, TTO valuations were demonstrated to be influenced by duration effects whereby the number of remaining years in a health state may determine the decision to trade-off (Attema & Brouwer, 2010; Sackett & Torrance, 1978). Another common issue is that of time preference whereby life years in the near and distant future are valued differently. Most people are found to have positive time preference, with time in the near future given greater preference than the distant future (van der Pol & Roux, 2005).

While DCE is able to establish the relative merit of one health state to another, these preference values lie on an ordinal scale from "best" to "worst" and need to be anchored on dead and full health if QALYs are to be derived from DCE tasks (Rowen, Brazier, & van Hout, 2015).

1.5 EQ-5D in Malaysia

The EQ-5D exist for the adult population use in two forms. The three level (EQ-5D-3L) version and the newer five-level version (EQ-5D-5L) with reduced ceiling effects and improved sensitivity (Devlin & Brooks, 2017; Janssen, Birnie, Haagsma, & Bonsel, 2008; Janssen, Pickard, et al., 2013).

The EQ-5D-3L has been validated for use in Malaysians and a value set was established in 2012 (Md Yusof, Goh, & Azmi, 2012; Shafie, 2014). However, due to the small sample size employed and the general lack of representativeness in the sampling applied, the value set is seldom used to inform CUA involving Malaysians (Shafie, 2014). On the other hand, prior to the undertaking of this study, the newer EQ-5D-5L has never been validated and no value set has been estimated for use in Malaysians.

1.6 Health technology assessment in Malaysia

Malaysia's public healthcare system is funded by government general taxation revenue with the Ministry of Health Malaysia (MOH) being the primary healthcare services provider. The continual rise of healthcare-related expenditures makes prudent resource allocation on a finite budget even more important.

As such, the use of health technology assessments (HTA) provides a strong incentive for prioritizing healthcare decision making. Collectively, HTA describe the process of systematic assessment of the properties and effects of health technologies, be it pharmaceuticals, medical devices, and even medical procedures for the introduction, procurement, implementation, and reimbursement decisions (Health Technology Assessment international, 2020).

There are two governmental bodies that utilize HTA in Malaysia, namely the Malaysian Health Technology Assessment Section (MaHTAS) of the Medical Development Division, and the Formulary Management Branch of Pharmacy Practice And Development Division (PPDD).

The activities of MaHTAS mainly include conducting full HTA, technology reviews, and information briefs to inform policy decisions relating to a variety of health technologies, besides developing evidence-based clinical practice guidelines and carrying out horizon scanning of emerging technologies ("MaHTAS – health technology assessment section, Ministry of Health Malaysia,"). On the other hand, the HTA scope of PPDD covers conducting technology reviews of pharmaceuticals exclusively for listing in the MOH Medicines Formulary (MOHMF) (Hussain, 2008; Shafie, Chandriah, Yong, & Wan Puteh, 2019). The MOHMF listing of drugs is necessary for the use of drugs in MOH facilities.

The HTA process generally considers the aspects of safety, effectiveness, and economic evidence in assisting decision making. Economic evidence facilitates explicit and transparent cost and consequence comparison especially important in informing a limited budget (Drummond et al., 2015b). According to the Malaysian Pharmacoeconomic Guidelines, QALYs is the preferred outcomes measure when HRQoL is an important outcome and when the intervention affects both morbidity and mortality. Outcomes should also be based on validated instruments and local preferences should inform CUA when such data are available (Pharmaceutical Services Division, 2012).

However, the current lack of local epidemiology and utility estimates hinders HTA to use CUA for healthcare prioritization in Malaysia, and was noted by Shafie and colleagues (Shafie, Chandriah, et al., 2019).

1.7 Problem statement

The current lack of a representative health preference value set limits the extensive use of CUA in informing healthcare decision making in Malaysia. The difference in sociodemographic characteristics have been shown to significantly influence health preference patterns and the proportion of ethnicities in Malaysia, especially, is quite dissimilar compared to her neighbouring countries. Such differences lead to different dimensions for healthcare prioritization, thus justifying the need for a country-specific value set. However, before an instrument can be applied to generate a value set, it should demonstrate certain psychometric properties for it to truly reflect the preference of the population. The EQ-5D-5L has never been validated in the Malaysian setting and no Malaysian-specific value sets are currently available. Comparability of a Malaysian EQ-5D-5L value set characteristics with value sets of other settings are also unknown. Additionally, investigations into the factors that affect these health preference patterns of Malaysians have never been conducted before.

1.8 Research questions

Therefore, this study would like to first address how the EQ-5D-5L questionnaire is generally perceived by Malaysians and could there be potential dimensions that can complement the current ones in the EQ-5D-5L. Thus, if present, the study wants to explore how these additional dimensions would affect the current EQ-5D-5L descriptive measure.

Subsequently, the performance of the EQ-5D-5L in term of psychometric properties in the Malaysian general population is still unknown and would be tested accordingly. The health preference of Malaysians elicited based on the EQ-5D-5L value set and the comparison with a Malaysian EQ-5D-3L would also be investigated. Lastly, the study would like to answer how sociodemographic factors and religiosity affect EQ-5D-5L valuation patterns in the Malaysian population.

CHAPTER 2

LITERATURE REVIEW

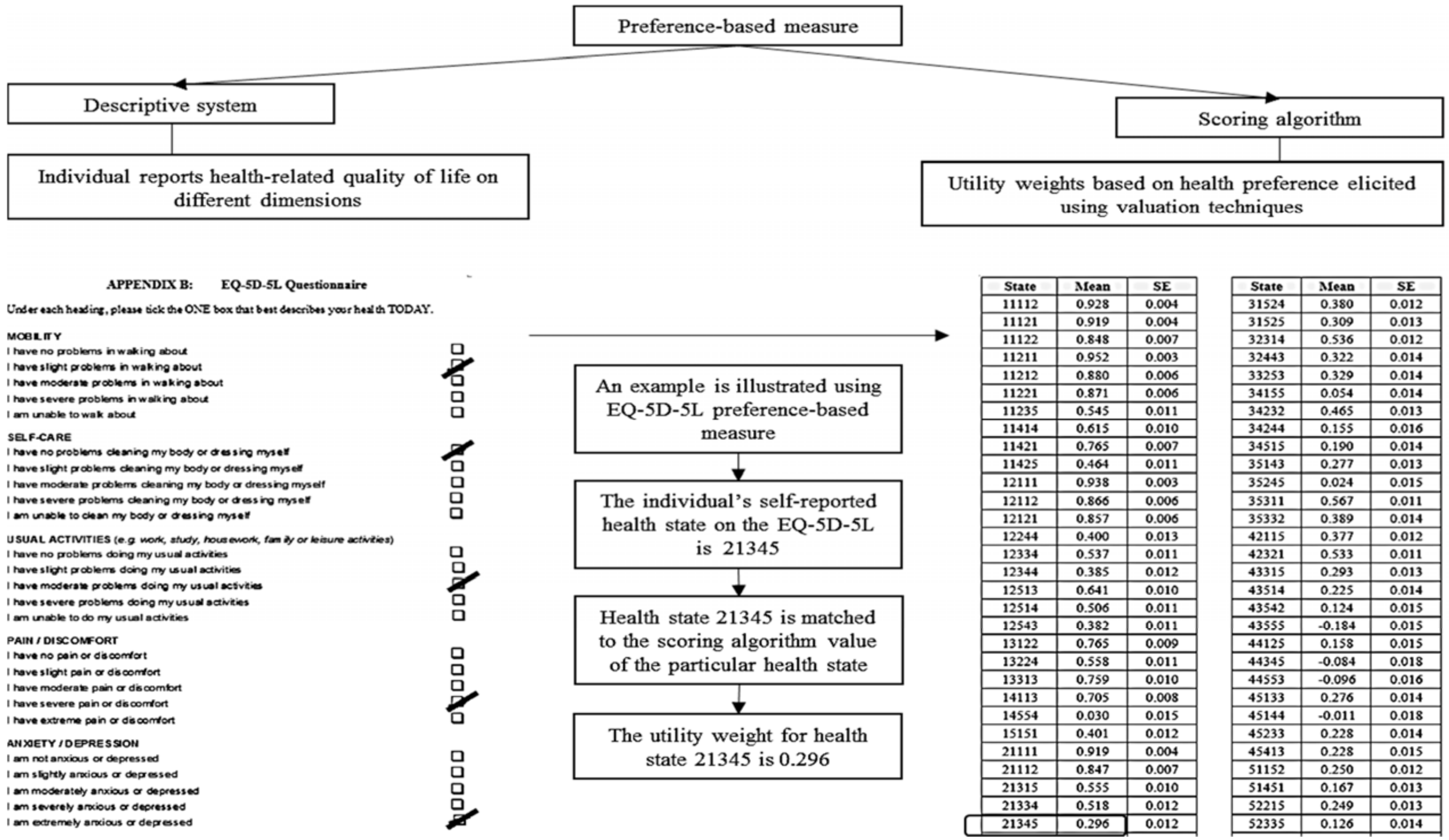
2.1 Generic preference-based measures (PBMs)

PBMs comprise two distinct parts, namely a standardized descriptive system for describing impact on HRQoL and a scoring algorithm to assign utility weights to each health state the system is able to describe (Figure 2.1). Valuations of the health states are usually conducted among the adult general population to inform the scoring algorithm for use, although those that cater for children are being actively developed by colleagues in Indonesia (Rowen, Rivero-Arias, Devlin, & Ratcliffe, 2020).

When applying these values, the patient is first asked to report their health on the descriptive system. Then, using the algorithm, single index scores are generated for the health state of the patient. Anchored on the scale where full health is equivalent to one and death is zero, these preference-based scores are then applied to calculate QALYs (Brazier & Ratcliffe, 2016).

PBMs can exist as generic or condition-specific (Goodwin & Green, 2016). Generic ones consist of dimensions representing a broader spectrum of health which can be applied across health conditions while dimensions of condition-specific measures are tailored to the specific disease (Versteegh, Leunis, Uyl-De Groot, & Stolk, 2012).

Generic PBMs are the most commonly used form of generating QALYs to inform economic evaluations mainly because they are easy to complete and comparability between patient groups is facilitated due to the generic and broad nature of the dimensions included (Brazier, Ara, Rowen, & Chevrou-Severac, 2017).



State	Mean	SE
11112	0.928	0.004
11121	0.919	0.004
11122	0.848	0.007
11211	0.952	0.003
11212	0.880	0.006
11221	0.871	0.006
11235	0.545	0.011
11414	0.615	0.010
11421	0.765	0.007
11425	0.464	0.011
12111	0.938	0.003
12112	0.866	0.006
12121	0.857	0.006
12244	0.400	0.013
12334	0.537	0.011
12344	0.385	0.012
12513	0.641	0.010
12514	0.506	0.011
12543	0.382	0.011
13122	0.765	0.009
13224	0.558	0.011
13313	0.759	0.010
14113	0.705	0.008
14554	0.030	0.015
15151	0.401	0.012
21111	0.919	0.004
21112	0.847	0.007
21315	0.555	0.010
21334	0.518	0.012
21345	0.296	0.012

State	Mean	SE
31524	0.380	0.012
31525	0.309	0.013
32314	0.536	0.012
32443	0.322	0.014
33253	0.329	0.014
34155	0.054	0.014
34232	0.465	0.013
34244	0.155	0.016
34515	0.190	0.014
35143	0.277	0.013
35245	0.024	0.015
35311	0.567	0.011
35332	0.389	0.014
42115	0.377	0.012
42321	0.533	0.011
43315	0.293	0.013
43514	0.225	0.014
43542	0.124	0.015
43555	-0.184	0.015
44125	0.158	0.015
44345	-0.084	0.018
44553	-0.096	0.016
45133	0.276	0.014
45144	-0.011	0.018
45233	0.228	0.014
45413	0.228	0.015
51152	0.250	0.012
51451	0.167	0.013
52215	0.249	0.013
52335	0.126	0.014

Figure 2.1 Parts of a preference-based measure

2.1.1 Quality of Well-Being Scale

The earliest of PBMs, the Quality of Well-Being Scale (QWB) was developed in the 1970s using theory from the General Health Policy Model to inform health services resource allocation (Fanshel & Bush, 1970; Kaplan & Anderson, 1996). While the original version of QWB was designed to be interviewer-administered, a newer self-administered version, QWB-SA includes additional symptom assessment (Kaplan & Anderson, 1988).

QWB-SA's descriptive system consist of two parts which are i) three function-related dimensions (mobility, physical activity, and social functioning) and ii) a list of symptoms/problems. The 46 functional levels (including death) and 68 complexes (from 27 in the interview-administered version) form a total of 945 health states. Preference weights for the QWB-SA were estimated from a sample of 435 adults of primary care clinics and college campuses in San Diego, USA using the VAS valuation technique (Seiber, Groessl, David, Ganiats, & Kaplan, 2008). These weights are universally used to inform the scoring algorithm. The QWB instrument and accompanying scoring algorithm are available for free for non-profit use from the developers (The UCSD Health Services Research Center, 2020).

2.1.2 Health Utilities Index

First developed in Canada in the late 1970s to inform an economic evaluation of neonatal intensive care, the Health Utilities Index mark 1 (HUI1) was replaced by two newer versions (Torrance, 1982). HUI2 is a generic PBM for children and was originally designed for applications in childhood cancer while HUI3 was developed for use in adults (Feeny et al., 2002; Torrance et al., 1996). The PBM can be self-completed, administered by an interviewer, or by a proxy.

HUI3 composes of eight dimensions which are vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain (Health Utilities Inc, 2020). Each dimension has four to five levels and with a total of 972,000 health states defined by the descriptive system. HUI3 was originally valued in Hamilton, Ontario by 504 adults, representative of the population using mainly VAS and a few states with both VAS and SG. A power function was then developed between the two valuation techniques to convert VAS values into SG. While preference weights have also been generated in a study in France by 365 general public respondents, the scoring algorithm of the Canadian one is generally applied in outcomes research (Le Galès, Buron, Costet, Rosman, & Slama, 2002). An administration fee is charged to those wanting to use the measure (Health Utilities Inc, 2020).

2.1.3 15D

The 15D was developed in Finland in the early 90s and the latest version is the 15D.2, which is recommended for use (Sintonen & Pekurinen, 1993; Sintonen & Richardson, 1994). The instrument is self-completed by the respondents.

The 15 dimensions in this PBM include mobility, vision, hearing, breathing, sleeping, eating, speech, elimination, usual activities, mental function, discomfort and symptoms, depression, distress, vitality, and sexual activities. Each dimension has four to six levels and consequently the number of health states described accounts to 31 billion.

The scoring algorithm was derived from valuation of five random Finnish samples, each consisting of 500 respondents and the valuation technique applied was

VAS (Sintonen, 1995). To use the measure, permission has to be obtained from the developer (Sintonen, 2020).

2.1.4 EQ-5D

The EQ-5D-5L, sometimes referred to as EuroQol five-dimensional questionnaire was developed by The EuroQol Group, consisting of multidisciplinary and multi-country-based researchers in the late 80s (Brooks, 1996). Besides the EQ-5D version for adult population use, a newer EQ-5D-Y is available for use in children (Devlin & Brooks, 2017). EQ-5D can be self-completed or administered via face-to-face interviews. It is reportedly the most widely used PBM available (Richardson, Mckie, & Bariola, 2014).

The five dimensions consist of mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Responses on the EQ-5D can be combined together in the above dimensional order to for a health state. The original version has three levels of severity (EQ-5D-3L) and a newer five level one (EQ-5D-5L) are available for use, defining a total of 243 (3^5) and 3125 (5^5) health states respectively. (Herdman et al., 2011). The questionnaire is composed of two sections, a descriptive system describing the five dimensions, and a 20-cm thermometer like scale known as the EuroQol Visual Analogue Scale (EQ-VAS) on which respondents rate their general health but is not needed to derive preference weights.

The first country-specific scoring algorithm was derived by the UK Measurement and Valuation of Health (MVH) group using 2997 general population samples in the UK producing mainly a TTO-based value set (Dolan, Gudex, Kind, & Williams, 1996). It was the most commonly used preference weights available until various other countries started developing their respective value sets to better reflect the health preference of the population (Mahlich et al., 2018; Xie et al., 2014). DCE techniques have been developed for use in valuing EQ-5D health states too (Oppe, Devlin, van Hout, Krabbe, & de Charro, 2014; Stolk et al., 2010). The questionnaire is free for academic and research use and is available from the developers (EuroQol Research Foundation, 2020).

2.1.5 SF-6D

SF-6D was derived from the profile-based health status questionnaire, 36-Item Short-Form Survey, SF-36 by researchers at the University of Sheffield, UK. Two versions of SF-6D currently exist, versions one (SF-36v2) and two (SF-12v2) being derived from SF-36 and SF-12 respectively (Brazier, Roberts, & Deverill, 2002; Brazier & Roberts, 2004). The PBM can be administered via face-to-face or self-reported forms.

Both versions measure six dimensions of health, namely physical functioning, role limitation, social functioning, pain, mental health, and vitality. Each dimension has four to six levels and the health classification system defines 18,000 and 7,500 health states for version one and two of the SF-6F respectively.

SF-6D was first valued by 836 members of the UK general public and preference was elicited using SG. A DCE-based value set was estimated in Australia (Norman et al., 2014). SF-6D value sets are also available for a number of other countries including Japan (Brazier et al., 2009), Hong Kong (Lam, Brazier, & McGhee, 2008), and Brazil (Cruz et al., 2011). Portugal (Ferreira, Ferreira, Pereira, Brazier, & Rowen, 2010), Spain (Méndez, Abellán Perpiñán, Sánchez Martínez, & Martínez Pérez, 2011).

However, it is not generally recommended to administer the SF-6D directly (The University of Sheffield, 2020b). Instead, the SF-36v2 or SF-12v2 should be administered and the algorithm will then be used to convert these scores to the ones on

SF-6D. While the SF-6D measure and scoring algorithms are free for non-commercial use, SF-36v2 and SF-12v2 are copyrighted (Optum Inc, 2020; The University of Sheffield, 2020a) .

2.1.6 Assessment of Quality of Life

The Assessment of Quality of Life (AQoL) instrument was developed in Australia by researchers at the University of Melbourne and Monash (Hawthorne, Richardson, & Osborne, 1999). With originally five major dimensions and 15 items, the 8-dimensional AQoL-8D is latest version.

AQoL-8D describes health on eight dimensions, each with three to eight items (total of 35 items) and the items have four to six levels. The main eight dimensions are independent living, pain, senses, mental health, happiness, coping and relationships and a total of 2.37 x 10^{23} health states are defined (Richardson, Sinha, Iezzi, & Khan, 2014).

A stratified sample of 670 general public and mental disorder patient in Australia was used in estimating the health preference values. Valuations involved mainly VAS and a few TTO tasks which then are transformed to TTO-based values. The instrument and scoring algorithm are available for free from the developers (AQoL, 2014).

2.1.7 Considerations when choosing a generic PBM

According to Brazier and colleagues, the most important factors to consider when selecting a generic PBM are practicality, measurement properties (validity, reliability, responsiveness), valuation technique applied and comparability between instruments