Tampere University

MIHIR GANDHI

Methodological Issues in Health State Valuation in the General and Patient Populations

Tampere University Dissertations 3

Tampere University Dissertations 3

MIHIR GANDHI

Methodological Issues in Health State Valuation in the General and Patient Populations

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty Council of Medicine and Life Sciences
of the University of Tampere,
for public discussion in the auditorium F114
of the Arvo building, Arvo Ylpön katu 34, Tampere,
on 22.02.2019, at 12 o'clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Medicine and Health Technology

Supervisor(s)	Adjunct Professor Yin Bun Cheung Tampere University Finland	Professor Julian Thumboo Duke-NUS Medical School Singapore
Pre-examiner(s)	Docent Anna-Marie Aalto University of Helsinki Finland	Professor Jan Abel Olsen University of Tromso Norway
Opponent(s)	Professor Feng Xie McMaster University Canada	
Custos	Professor Per Ashorn Tampere University Finland	

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Cover design: Roihu Inc.

ISBN 978-952-03-0984-8 (print)
ISBN 978-952-03-0985-5 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN:978-952-03-0985-5

PunaMusta Oy
Tampere 2019

To my most respected supervisors, teachers, and parents

ABSTRACT

Background: Utility values reflect people's preferences for health states. The development of a value set, which assigns utility values to all possible health states described by a health state utility instrument, involves research participants valuing a number of health states. Previous studies have shown that valuation of disease-specific health states can be affected by adding a disease label on their descriptions. However, the impact of a label on generic health states is not clear. Another long-debated issue is whether value sets should be based on patients' or the general population's preferences. This matters if there are systematic differences in their valuations, for which limited research has been done. Another important but rarely studied issue is the sample size determination for value set studies.

Aims: (i) To evaluate how the most severe health state (a generic health state) labelled as "All-worst" is valued by the general population, (ii) to compare values for hypothetical health states elicited from patients with chronic disease and from the general population, and (iii) to propose methods for determining the sample size for value set studies for the latest version of EQ-5D health state utility instrument.

Materials: Data from three studies, referred to as A, B and C, were used. The analyses for aims (i), (ii), and (iii) were based on study A, A & B, and C, respectively. All studies were cross-sectional, face-to-face surveys of health state valuation conducted in participants from Singapore. Study A involved valuation for SF-6D and EQ-5D health states using the visual analog method in the general population (n = 1034). Study B involved valuation for EQ-5D health states using the composite time trade-off method in general population (n = 175), heart diseases patients (n = 175), and cancer patients (n = 175). Study C involved valuation for EQ-5D using the composite time trade-off method in the general population (n = 1000).

Methods: An ordinary least-square (OLS) regression model was performed to compare the valuation score between the most severe health states labelled as "All-worst" and other health states, adjusting for health state descriptors. Similar model with addition of socio-demographic covariates was performed to compare the valuation score given by patient populations and the general population. Finally, I

proposed four approaches for determining the sample size for the EQ-5D value set studies. The first approach was based on acheiving the desired width of prediction interval for valuation scores using an OLS model. The second approach empirically determined a sample size based on mean absolute error in predicting valuation scores using empirical data. The last two approaches were based on assessing the statistical significance and estimating regression coefficients with a desired width of confidence interval in the OLS model.

Results: In Study A (general population sample), about 50% of participants were female, and 12% eldrely (above 60 years). About 43% of participants self-reported having chronic diseases (14% rheumatism, 13% hypertension, 9% diabetes, 4% heart disese, and 4% lung disesaes). Participants in the patient's groups were older compared to healthy participants. In Study B sample, patient populations had a higher proportion of elderly (53% in heart disease and 33% in cancer *vs* 17% in general population). The heart disease group had a lower proportion of female (34% in heart disease *vs* 58% in cancer and 53% in general population). Study C (general population sample) had about 53% female, and 17% elderly participants.

Using data from Study A, I found that the valuation of the health state labelled as "All-worst" was significantly lower than the value expected according to its descriptors. I compared the valuation of health states by participants with chronic diseases and a general population in Studies A and B. Values of mild and severe states elicited by patients with heart disease were significantly different from the general population. No such difference was found between any other patient groups and the general population. Finally, I developed and demonstrated the four approaches for determining the sample size for value set studies using required parameters estimated from Study C.

Conclusions: A labelling effect is plausible in value set studies, and therefore labelling health states should be avoided. Patients with chronic diseases may differ from the general population in the strength of their preferences for hypothetical health states. Cost-utility analysis based on utility values derived from the general population may not accurately evaluate healthcare interventions for certain type of diseases, such as heart diseases. Finally, my methods proposed for sample size determination can help to decide on an appropriate sample size for value set studies.

TIIVISTELMÄ

Taustaa: Ihmisten terveydentilan arvostus heijastuu hyötyarvoihin. Jos halutaan muodostaa arvojoukko, jossa nimetään hyötyarvot kaikille mahdollisille terveyskyselyssä kuvatuille terveydentiloille, tutkimukseen osallistujien on arvioitava erilaisia terveydentiloja. Aiempien tutkimusten mukaan sairauskohtaisten terveydentilojen arviointiin voidaan vaikuttaa lisäämällä kuvaukseen sairauden nimi. Nimeämisen vaikutus yleisiin terveydentiloihin on kuitenkin epäselvää. Lisäksi keskustelua on käyty pitkään siitä siitä, pitäisikö arvojoukkojen perustua vain potilaiden vai yleisemmin koko valtaväestön mielipiteisiin. Tämä on olennaista, jos arvioinneissa on järjestelmällisiä eroja, joita on tutkittu vasta vähän. Toinen tärkeä, joskin harvoin tutkittu, aihe on arvojoukkotutkimusten otoskoon määrittäminen.

Tavoitteet: 1) Tutkia, miten valtaväestö arvioi kaikista vakavinta, pahimmaksi nimettyä (yleistä) terveydentilaa, 2) vertailla kroonisista sairauksista kärsivien potilaiden ja valtaväestön hypoteettisten terveystilojen arviointeja, sekä 3) ehdottaa keinoja EQ-5D-elämänlaatumittaria hyödyntävien arvojoukkotutkimusten otoskoon määrittämiselle.

Materiaalit: Tutkimuksessa käytettiin aineistoja kolmesta tutkimuksesta, joihin viitataan tutkimuksina A, B ja C. Tavoitteiden 1, 2 ja 3 analyysit perustuivat järjestyksessä tutkimuksiin A, A+B ja C. Kaikki tutkimukset olivat kasvotusten suoritettuja poikittaistutkimuksia, joissa singaporelaiset osallistujat arvioivat terveydentilaansa. Tutkimuksessa A arvioitiin SF-6D- ja EQ-5D-mittarien mukaisia terveydentiloja visuaalisella analogisella metodilla valtaväestön keskuudessa (n = 1 034). Tutkimuksessa B arvioitiin valtaväestön (n = 175), sydänsairaiden (n = 175) ja syöpäpotilaiden (n = 175) EQ-5D-mittarin mukaisia terveydentiloja yhdistetyllä laatupainotettujen elinvuosien tutkimuksella. Tutkimuksessa C arvioitiin valtaväestön (n = 1 000) EQ-5D-mittarin mukaisia terveydentiloja yhdistetyllä laatupainotettujen elinvuosien tutkimuksella.

Menetelmät: Aineiston analyysissa käytettiin regressioanalyysiä pienimmän neliösumman menetelmällä (PNS) pahimman ja muiden terveydentilojen arviointien vertailemiseksi. Menetelmää mukautettiin terveydentilojen kuvaajien mukaan.

Potilasryhmien ja valtaväestön arvioiden vertailua varten suoritettiin vastaava analyysi, johon oli lisätty sosiaalisia ja demografisia taustamuuttujia. Lopuksi esitetään neljä lähestymistapaa EQ-5D-arvojoukkotutkimusten otoskoon määrittämiseksi. Ensimmäisessä lähestymistavassa pyritään saavuttamaan haluttu ennustusväli arvioinneille PNS-mallin avulla. Toisessa lähestymistavassa otoskoko määritetään empiirisesti perustuen keskimääräiseen absoluuttiseen virheeseen arviointien ennustamisessa empiiristen tulosten perusteella. Kolmas ja neljäs lähestymistapa perustuvat tilastollisen merkitsevyyden ja regressiokertointen arvioinnille sekä PNS-mallin halutulle luottamusvälille.

Tulokset: Tutkimuksessa A väestöotoksesta noin 50 % vastaajista oli naisia ja 12 % yli 60-vuotiaita. Noin 43 % vastaajista ilmoitti kärsivänsä kroonisista sairauksista (14 % reuma, 13 % kohonnut verenpaine, 9 % diabetes, 4 % sydänsairaus, 4 % keuhkosairaus). Potilasryhmien vastaajat olivat terveitä vastaajia vanhempia. Tutkimuksen B otoksesta potilasryhmissä oli enemmän ikääntyneitä ihmisiä kuin valtaväestössä (sydänsairaat: 53 %, syöpäpotilaat: 33 %, valtaväestö: 17 %). Sydänsairaiden ryhmässä naisten osuus oli pienempi kuin muissa ryhmissä (34 %, vs. 58 % syöpäsairaissa ja 53 % valtaväestössä). Tutkimuksen C väestöotoksesta noin 53 % oli naisia ja 17 % ikääntyneitä.

Tutkimuksen A tuloksista kävi ilmi, että kaikista huonoimmaksi nimetyn terveydentilan arviointi oli merkittävästi matalampi kuin mitä oli odotettavissa kuvaajien perusteella. Tässä tutkimuksessa verrattiin tutkimusten A ja B kroonisesti sairaiden vastaajien terveydentila-arvioita samojen tutkimusten väestöotoksiin. Sydänsairaiden vastaajien lievien ja vakavien tilojen vastaukset poikkesivat merkittävästi valtaväestön vastauksista. Tällaista eroa ei löytynyt minkään muun potilasryhmän ja valtaväestön välillä. Lopuksi kehitettiin ja esiteltiin neljä tapaa määrittää arvojoukkotutkimuksen otoskoko tutkimuksen C pohjalta esitetyin parametrein.

Johtopäätökset: Sairauden tai terveydentilan nimeäminen saattaa vaikuttaa arvojoukkotutkimuksen tuloksiin, joten sitä tulisi välttää. Kroonisesti sairaiden potilaiden vastaukset saattavat poiketa valtaväestöstä hypoteettisten terveydentilojen suosimisen voimakkuudessa. Valtaväestön vastauksista saatujen hyötyarvojen kustannus-hyötyanalyysin avulla ei välttämättä voida arvioida terveydenhuollon toimia tarkasti joidenkin sairaustyyppien, kuten sydänsairauksien, osalta. Otoskoon määrittämiseen ehdottamani tavat saattavat auttaa sopivan otoskoon valitsemisessa arvojoukkotutkimuksiin.

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

The thesis is based on the original articles as below which are referred to in the thesis by the roman numerals:

I. Gandhi M, Thumboo J, Wee HL, Luo N, Cheung YB. How is the most severe health state being valued by the general population? *Health and Quality of Life Outcomes*. 2014; 12: 161. doi: 10.1186/s12955-014-0161-9.

II. Gandhi M, Thumboo J, Luo N, Wee HL, Cheung YB. Do chronic disease patients value generic health states differently from individuals with no chronic disease? A case of a multicultural Asian population. *Health and Quality of Life Outcomes*. 2015; 13: 8. doi: 10.1186/s12955-014-0200-6.

III. Gandhi M, Tan RS, Ng R, Choo SP, Chia WK, Toh CK, Lam C, Lee PT, Latt NKZ, Rand-Hendriksen K, Cheung YB, Luo N. Comparison of health state values derived from patients and individuals from the general population. *Quality of Life Research*. 2017; 26(12): 3353-3363. doi: 10.1007/s11136-017-1683-5.

IV. Gandhi M, Xu Y, Luo N, Cheung YB. Sample size determination for EQ-5D-5L value set studies. *Quality of Life Research*. 2017; 26(12): 3365-3376. doi: 10.1007/s11136-017-1685-3.

LIST OF TABLES AND FIGURES

TABLES

FIGURES

ABBREVIATIONS

CI	Confidence Interval
CUA	Cost-Utility Analysis
EQ-5D	EuroQol – 5 Dimensions
EQ-5D-3L	3-level EuroQol – 5 Dimensions
EQ-5D-5L	5-level EuroQol – 5 Dimensions
EQ-VT	EuroQol – Valuation Technology
HTA	Health Technology Assessment
MAE	Mean Absolute Error
MID	Minimal important difference
NCDP	No Chronic Disease Population
OLS	Ordinary Least-Square
p	P-value
PI	Prediction Interval
QALY	Quality-Adjusted Life-Year
QoL	Quality of Life
RE	Random Effects
SD	Standard Deviation
SE	Standard Error
SF-6D	Short Form – 6 Dimension
TTO	Time Trade-Off
VAS	Visual Analogue Scale

1 INTRODUCTION

Patients with chronic diseases and their families do not single-mindedly pursue increased survival when deciding between treatments. They also take into consideration other factors such as the cost of care and quality of life (QoL).(Malhotra et al., 2015) Hence, a QoL value in conjunction with cost and improvement in survival is important for choosing fairly between different treatments. Cost-utility analysis (CUA) is a special form of cost-effectiveness analysis used for the health technology assessment (HTA) in which the result is presented in cost per quality-adjusted life year (QALY) gained. QALY is the product of the time spent in a health state and its preference (also called utility). A treatment that requires less cost per QALY gained is considered more cost-effective.

CUA requires a value set having utility values of all possible health states according to a health state utility instruments, such as EuroQol-5Dimension (EQ-5D) and Short Form-6Dimension (SF-6D). It is developed by surveying in a patient population or general population which asks participants to rate their preferences for different health states, including the most severe health state and dead state, presented to them. The preference rating can be performed in various ways such as visual analog scale (VAS), time trade-off (TTO) and standard gamble (SG) methods.(Xie et al., 2014) A higher utility value indicates the health state is preferred.

Despite past successes in developing and applying CUA methods and its components, various methodological issues and uncertainties have remained. They can hinder the accuracy and appropriateness of CUA and methodological development for health outcome valuation.

One of the key concerns during the health state valuation is – how to minimize impact of external factors including survey research procedures so that the derived utility value reflects participant's true preference for the given health state, but not the artifact effect of external factors. For example, it is not uncommon to label the most severe health state as 'all worst' or 'pits' in addition to its description in value

set studies.(Ratcliffe et al., 2016, Craig et al., 2013, Wee et al., 2006) But the impact of such labeling on its value is unknown.

Another key concern is the sample size requirement for developing a value set, specifically for many low-resourced and/or small countries. However, methods for determining an appropriate sample size for value set studies is rarely studied. The sample size methods used in previous value set studies were either not aligned with the analysis strategy and/or not capitalized for the valuation protocol used. The majority of such sample size determination method studies were based on the landmark Measurement and Valuation of Health (MVH) study protocol for 3-level EQ-5D (EQ-5D-3L) utilizing the conventional TTO method.(Dolan, 1997) However, so far, no statistically justified sample size determination method has been proposed for any other valuation protocol, even for the recent standard protocol for 5-level EQ-5D (EQ-5D-5L) value set studies adopting the composite TTO method. Therefore, what should be the sample size for a value set study is an open question.

A long-debated issue that is increasingly getting attention is – how to incorporate patient preferences in the HTA.(Brazier et al., 2017a, Underwood, 2016, Bridges and Jones, 2007) Many countries across the world are promoting patient-centric approaches in healthcare management. This includes involvement of patients in the treatment decision. The question of whose preferences (utility), patient or the general population, should be used in the clinical decision making, and reimbursement for healthcare cost have not reached consensus.(Zhang et al., 2017, McTaggart-Cowan, 2011, Stamuli, 2011) Furthermore, it is not entirely clear whether patients have different preferences than the general population.(Ogorevc et al., 2017, Karimi et al., 2017a, Peeters and Stiggelbout, 2010) Limited research has been done in this direction. The issue is particularly important for countries where healthcare cost is not heavily subsidized by the government.

The issues mentioned above are just some of many issues to be resolved in the field of CUA and health state valuation. In this thesis, I investigated three methodological issues in health state valuation and provided recommendations to improve the valuation process. These issues are discussed in detail in the next chapter.

2 REVIEW OF LITERATURE

2.1 Approach to the Literature Review

The objective of the literature reivew is to summarize the previous research done related to the aims of my research presented in this thesis. Primarily, the review included literature on labeling effect on valuation of health states, impact of chronic disease experience on valuation of hypothetical health states, and sample size determination methods used in value set studies.

The literature search was performed for electronic publications indexed in PubMed. Publications were indentified using combination of key words such as health state valuation, general population, patient population, hypothetical health states, sample size, value set, labeling health state/health condition. Relevent publications were shortlisted from reviewing the publication titles and abstracts. Relevent references cited in the shortlisted publication were also included in the literature review. Additional literatue such as reports published by governments, and non-goverment or research organizations were included from known sources. The last search was performed on February 2018 with emphasize to include recent publications.

2.2 Health State Utility Instruments

A value set provides utility values for all possible health states described by a health state utility instrument. As value sets are generally used for CUA for comparing cost-benefit ratio of health products across different therapeutic areas, health states for value sets are usually described using generic (not specific to any disease) health state utility instruments. The most commonly used generic health state utility instruments are EuroQoL-5 dimension (EQ-5D) and Short Form-6 dimension (SF-6D).(Brazier et al., 2017a)

2.2.1 EQ-5D

The EQ-5D is the most widely used generic health state utility instruments for CUA.(Wisløff et al., 2014) It comprises of five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) plus a visual analog scale (EQ-VAS) of the overall health status. There are many language versions of the EQ-5D validated in various disease groups in many countries including Singapore.(Devlin and Brooks, 2017) It is the health state utility instrument (characterized by its descriptive system and valuation technique) preferred by the National Institute for Health and Care Excellence (NICE), United Kingdom, for eliciting utility values.(NICE, 2013) The original version of the EQ-5D (EQ-5D-3L) has 3 response levels (no problem, moderate problems, and extreme problems) for each of the five dimensions (see **Appendix 1**). EQ-5D-3L health states are defined by combining the five dimensions at different response levels. For example, the health state '11211' indicates no problem on any of the five dimensions, except 'moderate problems' with doing 'usual activities' (3rd dimension at response level 2). A total of 243 health states can be described using the EQ-5D-3L.

Recently, the EQ-5D's response levels have been revised from 3-level to 5-level (no problem, slight problems, moderate problems, severe problems, and extreme problems) (see **Appendix 2**). This revision has been shown to increase measurement precision. A total of 3125 health states can be described using the 5-level EQ-5D (EQ-5D-5L).

2.2.2 SF-6D

The Short Form-36 is 36 items health survey questionnaire used to assess patient health.(Ware Jr and Sherbourne, 1992) Short Form-36 has six dimensions – physical functioning, role limitations, social functioning, pain, mental health, and vitality. Each dimension has 4 to 6 levels. The Short Form-36 is available in many languages and also validated in several countries for a variety of disease populations.(Kwan et al., 2016, Thumboo et al., 2013, Thumboo et al., 2001) The SF-6D health state descriptive system represents one item from each of the 6 dimension of the SF-36 (see **Appendix 3**).(Brazier et al., 2002) Thus, a total of 18,000 health states can be described using the SF-6D.

2.3 Valuation Methods

Preference values elicited for health states are called utility or index values. A set of preference values elicited for all possible health states based on a given health state utility instrument is called a preference set, utility set, value set or index set. Utility values for different health states are determined by surveying patients or the general population which asks participants to rate their preferences for different health states presented to them. There are several valuation methods such as the VAS, TTO, standard gamble, and the like.(Brazier et al., 2017a, Gudex et al., 1996) Furthermore, there are several variants of these methods are also available and used to develop value sets. I used the VAS and TTO methods in my studies described in this thesis. The variants used in my studies are explained here.

2.3.1 Visual Analog Scale Method

This method involves two steps. In the first step, the participant is presented with a hypothetical scenario by showing 'immediate death' and the most severe health state (e.g., '33333' for EQ-5D-3L) and then asked whether he/she prefers to die now (immediate death) or live rest for the rest of life in the most severe health state. The state that the participant considered less desirable is assigned a value 0 on a 100-point VAS scale (similar to a thermometer) with two ends 100 (most desirable) and 0 (least desirable).(Gudex et al., 1996) **Figure 1** shows the VAS scale used in one of the studies (Study A described in section 4.2) used in this thesis. There could be possibility of using the EQ-VAS (Appendices 1 and 2) as the VAS scale. In the second step, the participant is asked to value a unique set of health states (e.g., 10 health states) from a pre-defined subset of all possible states (e.g., 243 health states for the EQ-5D-3L). In addition, the most severe or dead state, which one is not valued earlier at 0, is asked to value along with other states. The participant is asked imagining him/herself in the given health states for the rest of the life without changing and indicate health state positions for valuing health states on the VAS scale. Multiple health states can assign same value. Usually, the VAS valuation score is rescaled using an appropriate transformation function to get utility value representing 1 for perfect health, 0 for dead state, and negative values for worse than death states.

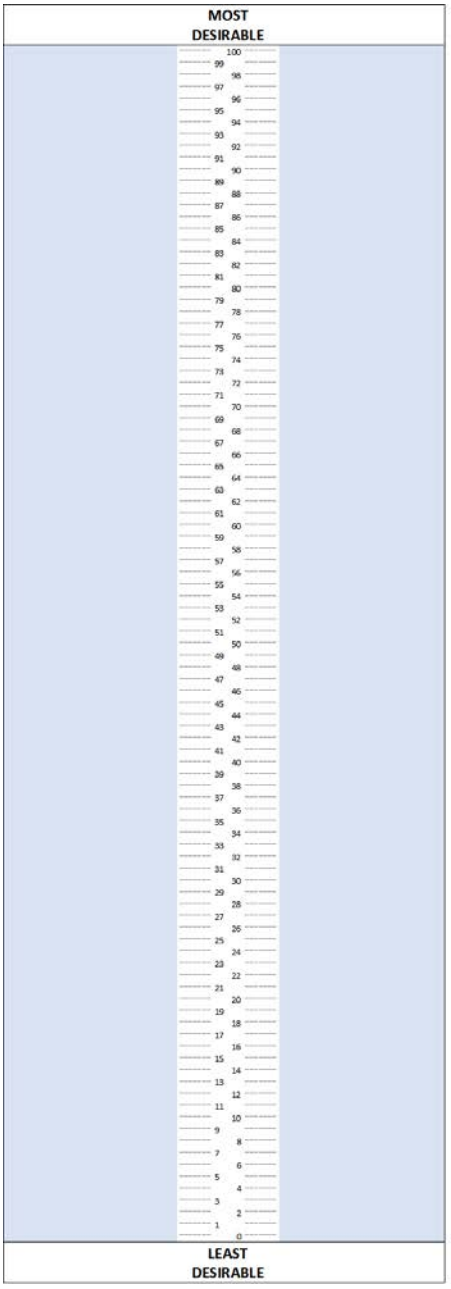


Figure 1. Visual analog scale for health state valuation

2.3.2 Time Trade-Off Method

Like the VAS method, the participant is asked to value a unique set of health states from a pre-defined subset of states in the TTO method. The TTO method involves an iterative procedure for each of the health states. The participant is asked how many years (x) in full health he/she would consider equivalent to a given amount of time (t) in an (unhealthy) health state.(Oppe et al., 2016, Oppe et al., 2014) The time t is generally set at 10 years. At the start, the number of years in full health is set at 10 and then reduced gradually to the point at which the participant is indifferent about the choice between the health state (e.g., '21121' EQ-5D-5L health state in **Figure 2**) in 10 years and full health for a shorter period (x). The utility of a health state is calculated as $x/10$.

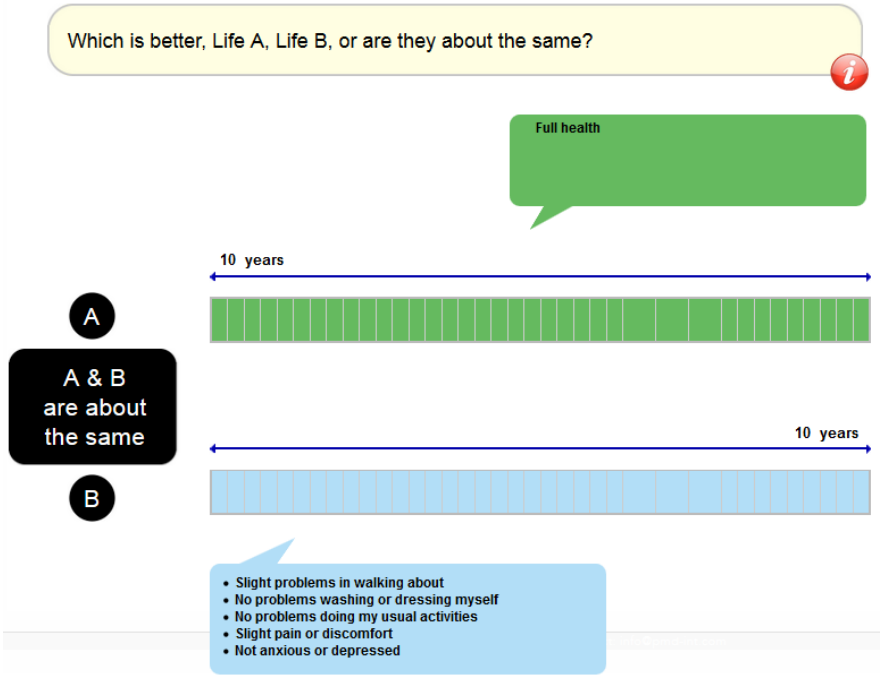


Figure 2. The time trade-off method with a 10-year time frame to value states worse than death

If a health state is very severe (e.g., '35554' EQ-5D-5L health state in **Figure 3**), the participant may consider experiencing this for 10 years so distasteful that he/she would consider it worse than death and prefer immediate death (i.e., $x \leq 0$). Such health states are valued using the lead-time TTO method.(Oppe et al., 2016, Oppe et al., 2014) In this approach, the health state is measured by asking the participant

to imagine a life of 20 years, in which the first 10 are in full health (i.e., lead time) and the remaining 10 in the health state, and then searching for the indifference point between this life and x years of full health. The utility of such health states is calculated as (*x*-10)/10.

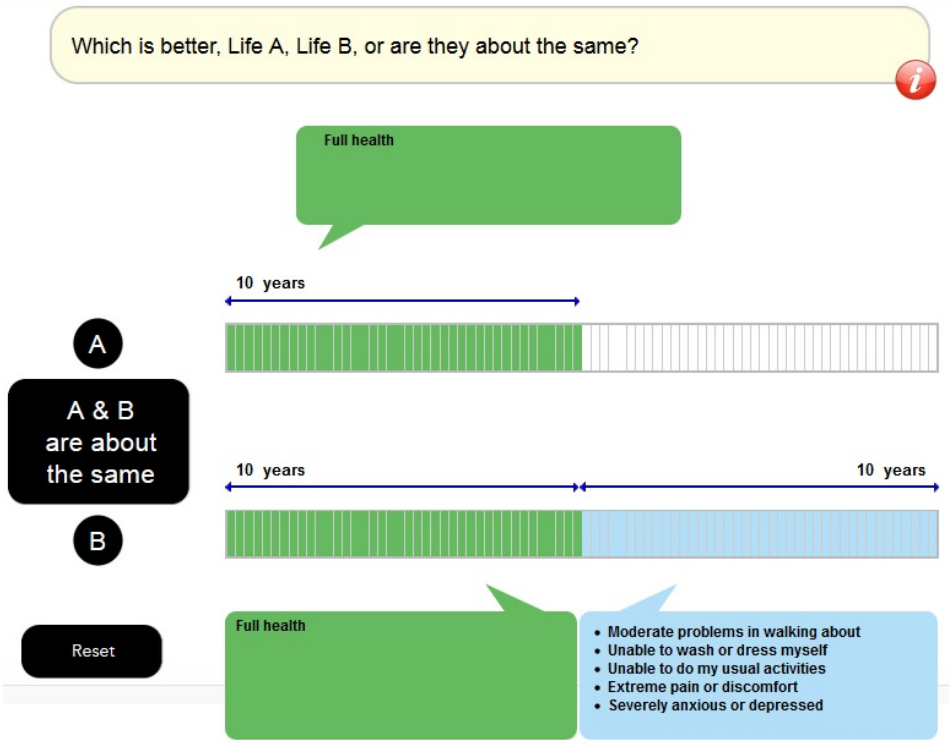


Figure 3. The lead-time time trade-off method with a 20-year time frame to value states worse than death

In the composite TTO method, the participant is first asked whether he/she considers the given health state better or worse than death. The health states better than death are valued using the conventional TTO (**Figure 2**), and the health states worse than death are valued using the lead-time TTO (**Figure 3**). The composite TTO method ensures that the utility value of each health state is bounded at -1 and 1, with 0 represents value for the 'dead' state. This approach is recommended by the EuroQol Research Foundation (the developer of EQ-5D) in the EuroQol Valuation Technology (EQ-VT) protocol for EQ-5D-5L value sets (explained later with more details).(Oppe et al., 2014)

2.4 Current Methodological Issues in Cost-Utility Analysis and Health State Valuation Methods

Over the last three decade methodologies for performing CUA and its components such as health state valuation are evolving. There is growing interest in evaluating health outcomes using QALY, and such approaches are adopted in policy guidelines by several regulatory agencies.(ISPOR, 2018) The research on developing health state utility instruments and valuation methodology assigning weights to health states/conditions described by such questionnaires has been a great success. It enabled to compare health benefits across different symptoms and diseases. Nevertheless, various methodological issues and uncertainties have remained, which can hamper the accuracy and appropriateness of CUA and health state valuation.

The widely used valuation methods such as the TTO and standard gamble, are complex, cognitively demanding, and time-consuming which make them difficult to implement in elderly and participants with low education level.(Karimi et al., 2017b, Brazier et al., 2017b) These methods are also vulnerable to external factors such as interviewer effect and survey research procedures. The newly emerging discrete choice experiment method is still in the exploratory stage, and several issues remain to be resolved before it can potentially replace the existing valuation methods.(Bansback et al., 2014)

In addition to these issues, sample size requirement for developing a value set is a major constraint for many low-resourced countries. However, methods for determining an appropriate sample size for value set studies is rarely studied. For example, the landmark Measurement and Valuation of Health (MVH) study in 1993 for EQ-5D-3L value set was conducted with a sample size of over 3000 participants from the United Kingdom.(Dolan, 1997) However, subsequently, a value set study in The Netherlands in late 2006 showed that a sample size of 300 participants could be sufficient.(Lamers et al., 2006) These sample size methods were based on the MVH study protocol for EQ-5D-3L value set studies using the conventional TTO method. The majority of sample size methods were also not aligned with the planned statistical analysis for developing the value set and/or capitalized on the valuation protocol. So far, no statistically justified method has been proposed for any other valuation protocol, even for the recent standard protocol for EQ-5D-5L value set studies. Therefore, what should be the sample size for a value set study is an open question. On another note, recently some research has been initiated to optimize

health states selection for direct valuation to improve the precision in predicting utility values and ultimately to reduce the number of health states to be valued and/or reduce the required sample size.(Yang et al., 2017) However, this research is based on the EQ-5D-3L health states using the VAS method. It will require being generalized for other valuation methods as well as health state utility instruments. Similarly, researchers in this field are actively trying different statistical models for enhancing the predictive ability and gain precision through modeling techniques, but so far limited success has been achieved.(Rand-Hendriksen et al., 2017, Ramos-Goñi et al., 2017)

Another long-debated issue which is increasingly getting attention is – how to incorporate patient preferences in the HTA.(Zhang et al., 2017, Underwood, 2016, Bridges and Jones, 2007) Many countries across the world are promoting patient-centric approaches in healthcare management. This includes involvement of patients in the treatment decision. The question of whose preferences (utility), patient or the general population, should be used in the clinical decision making, and reimbursement for healthcare cost have not reached consensus.(Brazier et al., 2017a, Stamuli, 2011, McTaggart-Cowan, 2011) Furthermore, it is not entirely clear whether patients have different preferences than the general population.(Ogorevc et al., 2017, Karimi et al., 2017a, Peeters and Stiggelbout, 2010) Limited research has been done in this direction. The issue is specifically important for countries where healthcare cost is not heavily subsidized by the government.

Along the line of incorporating patient preferences in HTAs, CUAs based on generic health state utility instruments may also be considered suboptimal for certain patient populations in which an essential dimension of QoL affected by the disease or its treatment is not adequately incorporated. For example, the EQ-5D lacks dimension related to vision, and speech. So, a CUA based on the EQ-5D utility values for evaluating treatment options in patients with cataracts or patients with stroke may be criticized. Therefore, research on adding extra items as "bolt-on" in the EQ-5D and developing its value set will be needed.(Yang et al., 2015) Consequently, the issue of comparability of such value sets with the existing value sets requires being addressed.

The issues mentioned above are just some of many issues to be resolved in the field of CUA and health state valuation. In this thesis, I investigated three methodological issues in health state valuation and provided recommendations to

improve the valuation process. These issues are discussed in detail in the subsequent sections.

2.4.1 Issue with Labeling of the Most Severe Health State

The most severe health state (e.g., '33333' for EQ-5D-3L and '645655' for SF-6D) described in a health state utility instrument is an important health state in the valuation process. Usually, either the most severe health state or dead state is valued lowest among all health states in valuation studies.(Green et al., 2000) The lower bound of utility values is therefore determined by them. Some previous studies have shown that valuation of severe health states can be affected by the survey research procedure and that their utility values are difficult to predict.(Al Sayah et al., 2016, Luo et al., 2007, Wee et al., 2006, Brazier et al., 2002)

Among the other factors that can affect the valuation of the most severe health state, special attention is required as to how its value can be affected if a label is added in its description. It is not uncommon to label the most severe health state as 'all worst' or 'pits' in value set studies.(Ratcliffe et al., 2016, Craig et al., 2013, Wee et al., 2006) There is an opinion that adding a disease label in disease-specific health state descriptions can have an impact on health state value, possibly due to prior knowledge or preconception of the disease.(Green et al., 2017, Robinson and Bryan, 2013) However, no study to our knowledge has investigated the impact of labeling a generic health state. Therefore, how the most severe health state (in the presence of a label) is valued needs attention.

2.4.2 Issue with Valuation by the General Population or a Patient Population

Panel on Cost-Effectiveness in Health and Medicine in the United States and the National Institute for Health and Care Excellence in England and Wales recommend using value sets elicited by a general population for CUA.(Sanders et al., 2016, NICE, 2013) A general population-derived utility is desirable when the utility is needed to inform decisions that allocate societal (taxpayer) resources.(Stamuli, 2011) However, a value set elicited by patients is preferred in clinical decision making from a patient-centered perspective. If the treatment costs are mostly paid by patients themselves, the patient-derived utility is also relevant for CUA for the comparison of treatment

within the given patient population. A systematic review reported that less than one-third of published CUAs use a general-population-derived utility; the remaining CUAs use a patient, clinician- or expert-derived utility, or authors' judgments.(Brauer et al., 2006) Patient-derived utility values are believed to be more accurate as patients who have experienced the disease conditions can appraise their condition(s) more accurately than individuals who have not experienced such conditions.(Karimi et al., 2017a, Russell et al., 1996) A meta-analysis reported that patients give higher values to their own health state than the individuals without such experience.(Peeters and Stiggelbout, 2010) However, there is conflicting evidence on whether patients with chronic diseases value hypothetical health states differently than individuals without chronic disease experience.(Ogorevc et al., 2017, Wang et al., 2014, Pickard et al., 2013, Krabbe et al., 2011) The conflicting results could be because the patients might have adapted to their condition or because individuals with no disease experience overestimate the impact of disease or disability on QoL.(Wang et al., 2014) Other factors such as the type of condition, the severity of the condition, and valuation method may also contribute to the differences in valuation. For example, a study conducted by Suarez-Almazor *et al.*(Suarez-Almazor and Conner-Spady, 2001) showed that individuals with arthritis value health states similar to individuals without arthritis when the VAS method is used. However, individuals with and without arthritis gave different valuations when the TTO method is used. It should also be noted that most of the studies evaluating differences in valuations between the chronic disease patients and the general population were conducted in the western population, with only one study being conducted in Asian population.(Ogorevc et al., 2017, Wang et al., 2014, Pickard et al., 2013, Krabbe et al., 2011, Peeters and Stiggelbout, 2010) Furthermore, the majority of the studies have compared valuation of patients own health state being experienced at the time of survey with hypothetical health states. Several studies have shown that cultural differences affect the valuation.(Luo et al., 2014, Johnson et al., 2005) There are several Asian countries including Singapore, China, and India where the majority of healthcare cost is paid by patients themselves.(WHO, 2014) In such scenarios, it is important to evaluate CUA using patients' 'experienced-based' value set. However, developing 'experienced-based' value set is challenging due to several reasons. It is difficult to ask patients who are experiencing very severe health states to provide value of their own health states. In some cases, it could be even impossible for health states such as '55555' for the EQ-5D-5L. Therefore, usually, the sample size for valuation of severe health states are limited and their validity could be sub-optimal. Furthermore, the patient own health should also be valued using a disease-specific

health state utility instrument to describe the current health state accurately. However, in reality, it is uncommon to have such validate instrument that can be used for valuation. An alternative is to use a generic health state utility instrument and ask patients to value hypothetical health states from their own perspective. It is likely that patients have experience more of such health states than the general population. So far there is little evidence is available about differences in hypothetical health state valuation between patients with chronic diseases and the general population to decide on whose value set should be used.

The question of whether patients and the general population have similar preferences for hypothetical health states has important implications on the use of the value set. If the answer is 'yes,' the use of utility values which reflects the health state preferences of the general population to inform medical decision making should be encouraged because they also reflect patients' preferences. If the answer is 'no,' caution should be exercised when using utility values, based on general population preferences to evaluate the treatment benefit of clinical interventions or quality of care. Nevertheless, the answer to this question is still elusive given the seemingly contradicting findings in previous studies.

2.4.3 Issue with Sample Size Determination of the EQ-5D-5L Value Set Studies

With the revision of the response levels from 3-levels to 5-levels in the EQ-5D, a new value set for the EQ-5D-5L health states is needed. The EQ-5D-3L value set studies were either mostly conducted using the VAS or TTO method with no standard protocol.(Szende et al., 2007) Therefore, value sets developed for different countries with different valuation protocols were not comparable. Nevertheless, the previous value set studies provided information on pros and cons of different variants of these two valuation methods. After conducting a few pilot studies to optimize the valuation method, the EuroQol Research Foundation proposed a revised TTO method, called composite TTO method for EQ-5D-5L value sets. The protocol is named the EuroQol Valuation Technology (EQ-VT) protocol (Oppe et al., 2014, Oppe et al., 2016) and considered the 'standard' for uniformly developing EQ-5D-5L value sets. It has been used for developing EQ-5D-5L value sets for several countries including England (Devlin et al., 2017), China (Luo et al., 2017), Canada (Xie et al., 2016), Netherlands (Versteegh et al., 2016), Germany (Ludwig et

al., 2018), Japan (Ikeda et al., 2015), Korea (Kim et al., 2016), Indonesia (Purba et al., 2017), and Singapore.

This standard protocol recommends a sample size of 1000 participants for a value set.(Oppe et al., 2014) The recommendation is based on some assumptions, such as 10 valuations per health state is sufficient, with limited theoretical justification. Furthermore, there is no empirical evaluation performed based on the data from the EQ-5D-5L value set studies on the appropriateness of the sample size. The selection of an optimal sample size is very important from both statistical and feasibility points of view. A study with insufficient sampler size lacks precision, and the use of unnecessarily large sample size leads to waste of resources and time. Although CUA is getting popularity and evidence of cultural differences in valuation is growing, many countries do not have country-specific value sets, possibly due to resource constraint for conducting a large-scale value set study. If a method allowing to estimate the sample size for desired precision or evaluating the impact of a sample size on the precision of predicted utility values is available, it will be very useful for planning value set studies according to study team's requirements.

3 AIMS

The research aims to provide recommendations for improving valuation techniques, development of value sets as well as the use of value sets. The specific aims were:

1. To evaluate how the most severe health state is valued by the general population.

2. To compare values for hypothetical health states elicited by patients with chronic disease and the general population. This has two sub-aims:

    2.1 To compare values for EQ-5D-3L health states elicited using the VAS method by individuals with self-reported chronic diseases (diabetes, rheumatism, hypertension, heart disease, lung disease) and by individuals with no chronic disease in the general population.

    2.2 To compare values for EQ-5D-5L health states elicited using the composite TTO method by individuals who were clinically diagnosed with chronic diseases (heart disease and cancer) and by individuals from the general population.

3. To propose methods for estimating the sample size for EQ-5D-5L value set studies according to the standardized procedure of the EQ-VT protocol.

4 METHODS

Data from three studies, referred to as A, B and C in this thesis for brevity, were used to achieve the research aims. The analyses for aims 1 and 2.1 were based on study A data. This study data was already available prior to the start of this thesis research. Study B was conducted to achieve aim 2.2. Study C data was used to demonstrate the methods proposed in aim 3. All three studies were conducted in Singapore.

4.1 Demography of Singapore

Singapore is a sovereign city-state in Southeast Asia with a land area of 719 km^2. It is ranked highly in GDP per capita (USD 52961), healthcare, life expectancy, quality of life, personal safety, housing and education.(Wikipedia, 2017) Singapore comprises of 5.5 million residents with 74.3% of Chinese descent, 13.4% Malay, 9.1% Indian, and 3.2% of other descent.(DOS, 2010) English is the most common language (80% literacy). Due to the scarcity of land, more than 80% of residents live in subsidized, high-rise, housing apartments (called "HDB" flats) built by the Housing and Development Board (HDB) of the government. **Table 1** shows some key sociodemographic and health indicators for Singapore.

Table 1. Basic sociodemographic and health indicators for Singapore

Indicator	Measure
Total area (km^2)	719
Total population (million)	5.54
Population density (per km^2)	7697
Population annual growth rate (%)	1.2
Average household size (persons)	3.39
Per capita gross domestic product (USD)	52961
Median household monthly income (USD)	6367
Literacy rate (%)	99.5
Life expectancy (years)	82.7
Population aged 65 years and above (%)	11.8
Ethnicity (%)	
Chinese	74.3
Malays	13.4
Indians	9.1
Others	3.2
Language most frequently spoken at home (%)	
English	62.4
Mandarin & Chinese dialects	4.7
Malay	4.3
Tamil	0.1
Others	28.6
Distribution of disability-adjusted life years (%)	
Cardiovascular diseases	20
Cancer	19
Neurological, vision and hearing disorders	14
Diabetes mellitus	10
Others	37

Source: Singapore Census 2010 (DOS, 2010) and Burden of Disease Study 2010 (MOH, 2010)

4.2 Study A: EQ-5D-3L and SF-6D Valuation Study in the Singapore General Population

4.2.1 Study Design and Participants

A cross-sectional, face-to-face survey of health state valuation for SF-6D and EQ-5D-3L using the VAS method was conducted in 2009 from a representative sample of the general population of Singapore. A multi-stage quota sampling approach was used to randomly select residential blocks, within which households were selected, within which one person per household was selected. Quota for ethnicity (400 Chinese, 400 Malay, and 234 Indian), gender (50% Female) and age (30% for 21–34 years, 40% for 35–49 years, 30% for 50+ years) set. Half of the participants within each ethnicity were interviewed in English and the remaining half in their native languages, i.e., Mandarin for Chinese, Malay for Malays and Tamil for Indians.

4.2.2 Valuation Procedure

Two separate samples of 1034 participants each were selected for the SF-6D and EQ-5D-3L value sets using the VAS method (explained in section 1.3.1). A subset of 249 SF-6D states (**Appendix 4**) was selected (out of 18,000) based on the protocol of Brazier *et al.*(Brazier et al., 2002) Each participant was first presented to a hypothetical scenario by showing 'immediate death,' and the most severe health states (e.g. '645655' for SF-6D) and then asked whether he/she prefers to die now (immediate death) or live rest for the rest of life in the most severe health state. Then, the participant was asked to value a unique set of 6 states from the subset of 249 SF-6D states and either dead or the most severe health state, depending on which one was not valued earlier at 0. The unique set of 6 health states were assigned to each participant in a way that they include health states of varying severity from mild to severe. The valuation of EQ-5D-3L health states was carried out similarly as SF-6D. A subset of 42 EQ-5D-3L health states (**Appendix 5**) was selected based on the protocol of Dolan.(Dolan, 1997) The most severe health states ('645655' for SF-6D and '33333' for EQ-5D-3L) were labeled as 'all-worst.' Unconscious state was also valued in addition to the other 6 assigned states.

4.2.3 Other Details

The study collected information about self-reported chronic diseases. The data collection form had a pre-defined list of chronic diseases which included diabetes, high blood pressure/hypertension, heart diseases, stroke, asthma or other lung diseases, cancer, rheumatism/back pain or other bone or muscle illness, mental illness (e.g., depression, anxiety neurosis, schizophrenia) and other illness (e.g., kidney problems or dialysis).

The study was approved by SingHealth Centralized Institutional Review Board.

4.3 Study B: EQ-5D-5L Valuation Study in the Singapore General and Patient Populations

4.3.1 Study Design and Participants

A cross-sectional, face-to-face survey of health state valuation for the EQ-5D-5L using the composite TTO method was conducted in 2016 in Singapore. Heart disease and cancer patients attending outpatient clinics at the National Heart Centre Singapore and the National Cancer Centre Singapore, respectively, were invited to participate in their routine visits. The National Heart Centre Singapore and the National Cancer Centre Singapore are the largest capacity specialty centers in Singapore for heart disease and cancer patients, respectively. The main eligibility criterion for the heart disease patients was hospitalization primarily for heart disease in the last 5 years; this criterion was used to screen out patients with mild heart conditions. The primary eligibility criterion for the cancer patients was to have histologically confirmed cancer of any type and stage in the last 5 years; this criterion was used to screen out cancer survivors. The study also included a sample from the general population recruited from three shopping malls. Shopping malls in Singapore (including ones used in this study) are located in proximity to local bus and train stations as well as food centers, which are part of the general population's daily routines. According to a lifestyle survey conducted by the Urban Redevelopment Authority of Singapore(URA, 2009), shopping malls are the third most regularly used facilities in Singapore. Therefore, sampling from shopping malls in Singapore is a good approximation of the general population. All participants were between 21 and

80 years old, able to read and communicate in English or Chinese (Mandarin), and well enough for an interview. A quota sampling based on age and gender distributions similar to the Singapore census was used to generate the general population sample.

4.3.2 Valuation Procedure

All interviews were conducted by the same trained interviewer using a computer program designed for the composite TTO method (see section 2.3.2) according to the EQ-VT protocol.(Oppe et al., 2014) Interviews were conducted in either English or Chinese according to participant's preference. The interviews with the patients were conducted in the hospitals in a quiet waiting area. The general population participants were interviewed in a quiet place of the malls where they were recruited.

The interviews started with some warm-up questions asking the participants to describe their own health using the EQ-5D-5L questionnaire. Subsequently, the composite TTO-based valuation task was explained to participants using the state of "in a wheelchair" as an example, after which three practice EQ-5D-5L health states were administered to familiarize participants with the task and EQ-5D-5L health states of varying severity. The practice states were followed by the composite TTO valuation of 10 EQ-5D-5L health states. The interviews ended with some feedback and background questions. A detailed description of the EQ-VT protocol can be found elsewhere.(Oppe et al., 2014)

In this study, all participants were asked to value the same set of 10 health states, in random order. The 10 health states were 11122, 21121, 21222, 21232, 32232, 32333, 22224, 31242, 53343, and 33453.

4.3.3 Other Details

In addition to the valuation interview, all the clinical information was collected directly from patients' medical records except information on functional classifications/status (listed below) were evaluated from patients by the interviewer. Clinical information included diagnoses, year of diagnosis, and clinical assessments such as New York Heart Association (NYHA) functional classification and Canadian Cardiovascular Society (CCS) functional classification of angina for heart disease

patients; and cancer stage and Eastern Cooperative Oncology Group (ECOG) performance status for cancer patients. All participants were also asked to self-report their current and past chronic diseases.

The study was approved by the SingHealth Centralized Institutional Review Board.

4.4 Study C: EQ-5D-5L Valuation Study in the Singapore General Population

4.4.1 Study Design and Participants

A cross-sectional, face-to-face national survey of health state valuation for EQ-5D-5L using the composite TTO method was conducted in 2016 in Singapore general population. A multi-stage quota sampling approach was used to randomly select residential blocks, within which households were selected, within which one person per household was selected. Sampling quotas were used to make the resultant sample resembling the general adult Singaporean population in distributions of age, gender, and ethnic groups. The eligibility criteria were aged 21 years or above; able to communicate and read in English, Chinese, or Malay.

4.4.2 Valuation Procedure

All interviews were conducted by trained interviewers using a computer program designed for the composite TTO method (EQ-VT protocol) similar to the one used in Study B. Interviews were conducted in either English, Chinese or Malay according to participant's preference. The interviews also included warm-up questions asking the participants to describe their own health using the EQ-5D-5L questionnaire, practice questions using wheelchair examples followed by composite TTO valuation of 10 EQ-5D-5L health states.

According to the EQ-VT protocol, each participant values a randomly allocated set (called a block) of 10 EQ-5D-5L health states.(Oppe et al., 2014) Each block includes one very mild health state from five pre-specified ones (21111, 12111, 11211, 11121, 11112), the most severe health state (55555), and eight health states from 80 pre-specified health states (from the remaining 3119 possible health states)

(**Appendix 6**). The protocol contains a total of 10 unique blocks. Simple random sampling is used to select one of the 10 blocks for each participant; hence the equal probability of each block being chosen.

4.4.3 Other Details

The study was approved by the National University of Singapore's Institutional Review Board.

4.5 Statistical Considerations

4.5.1 Sample Size Considerations

The sample size for study A was 2068 participants (1034 for SF-6D, and 1034 for EQ-5D-3L). It was decided for developing EQ-5D-3L and SF-6D value sets. As aims 1 and 2.1 were secondary analyses of study data already available, the entire sample was used for the research aim 1 and 1034 participants of the EQ-5D-3L sample for the research aim 2.1.

The sample size for study B was decided based on the research aim 2.2. The sample size was 525 participants (175 heart disease patients, 175 cancer patients, and 175 participants from the general population). It was estimated, to detect a difference of 0.1 standardized effect size in mean utility value of the 10 elicited health states, from a group of patients and general population using a two-sided test for 5% Type-I error rate and 80% power, assuming 10% participants may provide a logically inconsistent valuation or do not complete the interviews.

The sample size for study C was decided according to the recommendation of the EQ-VT protocol to use 1000 participants for the EQ-5D-5L value set studies. The study data was used to demonstrate sample size methods proposed in the research aim 3.

4.5.2 Analysis Populations

For research aim 1:

All participants enrolled in the study A except those who met the following criteria: a) valued less than 3 health states, b) did not value dead or the 'all-worst' state, c) valued dead or the 'all-worst' state or unconscious state higher than all the other states, d) gave the same valuation score to all the health states, e) self-reported or rated by the interviewers as having a poor understanding of health states description or valuation tasks.

For research aim 2:

Research aim 2.1: All participants enrolled for the EQ-5D-3L value set in the study A except those who met the following criteria: a) valued less than 3 health states, b) did not value dead or the 'all-worst' state, c) valued dead or the 'all-worst' state or unconscious state higher than all the other health states, d) gave the same valuation score to all the health states, e) self-reported or rated by the interviewers as having a poor understanding of health states description or valuation tasks. Furthermore, participants with chronic diseases other than diabetes, high blood pressure/hypertension, heart diseases, asthma/lung diseases, rheumatism/back pain/other bone-muscle illness were excluded as the number of participants with other chronic diseases were small (<10).

For research aim 2.2: All participants enrolled in the study B except those who met the following criteria: a) gave the same utility value to all the health states, b) gave negative or zero utility value to all health states (i.e., considered all health states worse than or equal to death).

For research aim 3:

All participants enrolled in the study C.

4.5.3 Analyses

4.5.3.1 General Considerations

All P-values (*p*) were two-sided. A *p* less than 0.05 was considered statistically significant. All confidence intervals (CI) were at 95% level. Regression models involving more than one records from a participant used the Eicker-Huber-White robust standard error (SE) for cluster data for statistical inference.(Williams, 2000) Minimally important differences (MID) of 4 points for EQ-5D-3L and 3.3 points for SF-6D on the 100-point VAS were considered to be of practical significance.(Walters and Brazier, 2003, Wee et al., 2007, Luo et al., 2010) Similarly, a MID of 0.05 points for the EQ-5D-5L utility values was considered to be of practical significance.(McClure et al., 2017, Nolan et al., 2016) All the analyses were carried out using Stata/MP version 10.1 or 13.1 for Windows.

4.5.3.2 Analyses for Research Aim 1

The analysis used untransformed raw valuation score (range 0: least desirable to 100: most desirable) to avoid any impact of transformation on the utility values. So far there is no consensus on which transformation is the most appropriate.(Lamers, 2007)

As the selection of whether the most severe health state is worse or better than dead in the first step of the VAS method can potentially affect the valuation of health states, the impact of labeling of the most severe health was studied separately among participants who considered it better or worse than dead. Accordingly, the mean valuation scores were presented separately using line graphs for these two group of participants.

An ordinary least-square (OLS) regression model was performed for the valuation score with 10 indicator variables representing 2 severity levels in 5 dimensions of the EQ-5D-3L referencing severity level 1 (no problem) and an intercept. The model also included an indicator variable (N3) to take into account additional disutility when a severe problem (level 3) is reported on at least one dimension.(Szende et al., 2007) In addition to the above commonly used variables, the model included two indicator variables D1 and D2 and their interaction: D1

represented the participant who considered the 'all-worst' state worse than dead, and D2 represented the 'all-worst' state. This model helped to assess whether there was a deficit in the valuation score for the most severe state even after taking the descriptors (levels in each dimension) into account. It also assessed the potential impact of considering the most severe state worse than death on its valuation. Similar OLS model was used to study the SF-6D valuation score. For the variable N3, the severe level was defined as levels 4–6 for physical functioning, levels 3–4 for role limitation, level 4–5 for social functioning, mental health and vitality, and level 5–6 for pain.(Brazier et al., 2002)

Perfect health state was not included in the models, as it was assigned a fixed value of 100 on VAS. The dead and unconscious states were also excluded from the models as they did not represent any health states/dimensions of SF-6D or EQ-5D-3L.

4.5.3.3 Analysis for Research Aim 2

The analysis for the research aim 2.1 was based on study A using the untransformed VAS valuation score whereas the analysis for the research aim 2.2 was performed using the data from the study B based on the composite TTO utility values.

We defined the control group as the individual without any chronic disease in study A and participants from the general population (of whom some may have chronic diseases) in study B.

Key sociodemographic characteristics of the patient and control groups were compared using either the Fisher's exact test or two-sample independent t-test.

EQ-5D-3L health states (total 42 states) in study A were classified as 'severe' (28 states) if at least one dimension was at level 3, and the remaining states (42 – 28 = 14 states) were classified as 'non-severe.' Similarly, 10 EQ-5D-5L health states in study B were classified as 'mild' (3 states: 11122, 21121, 21222) if all dimensions at level 1 or 2, 'severe' (4 states: 22224, 31242, 53343, 33453) if at least one dimension was at level 4 or 5, and remaining states as 'moderate' (3 states: 21232, 32232, 32333).

A separate analysis was performed to compare the valuations by each of the patient groups with those of the control group. Each analysis included an OLS

regression model (say core model) for the comparison of the overall difference in valuation (including all the health states) between a patient and the control groups. As in study A, different health states were valued by the participants; the core model also included indicator variables for health state descriptors (10 variables for severity levels in 5 dimensions of the EQ-5D-3L). The comparisons of valuation between a patient and the control groups for the non-severe and severe health states were performed by including an interaction term between the indicator variable for severe health state (reference: non-severe health state) and the indicator variable for the specific patient group (reference: the control group) in the core model. On the other hand, as the same 10 health states were valued by all participants in study B, indicator variables for health state descriptors were not included in the core model. The comparisons of valuation between a patient and the control groups for the mild, moderate and severe health states were performed by fitting the core model separately for each of the three types of health states. Each of the models for study A as well for study B was adjusting for key sociodemographic characteristics - Ethnicity, gender, age, marital status, education level, religion, house type or household income, and employment status of accounting for differences in the sociodemographic characteristics between the patients and control groups.

An exploratory analysis was performed based on study B. The differences in utility values between the health states valued by the same person were compared between the patients and control groups. As the difference in utility values of two health states is usually used to approximate utility gained from transitions between the two health states,(Sonnenberg and Beck, 1993) this comparison helped us to assess whether the different sets of utility values would give similar estimates when used to determine utility gained from health state transitions. First, differences in utility were calculated separately for each of the patient groups (heart disease and cancer), and the control group for 45 pairs of health states, where each pair has one health state better than the other in at least one of five dimensions. Second, differences in the differences between a patient group and the control group were presented using a line graph and tested by OLS models. The graph also presented the differences in differences adjusted for participants' sociodemographic characteristics, calculated using the OLS models performed separately for each pair of health states.

4.5.3.4 Analysis for Research Aim 3

4.5.3.4.1 Statistical Models for Estimating Utility Values

A value set provides an algorithm for deriving a utility value for each health state. A value set study conducted using the EQ-VT protocol provides values for only 86 directly valued health states (**Appendix 6**). These values are used to estimate values for all 3125 health states of the EQ-5D-5L, by regression analysis relating the elicited values to the health state descriptors. There are various ways to achieve it using an appropriate statistical model such as an OLS model, OLS model with cluster-robust SE, random effects (RE) model, interval regression model, and so on.(Rand-Hendriksen et al., 2017, Feng et al., 2017, Luo et al., 2014) Currently, there is no consensus on which model is ideal for predicting utility values. Usually, a study team performs several models and decides upon a final model based on the models' performance for consistency, bias, precision, and parsimony.

As an initial choice, an OLS model for estimating utility values can be specified as:

$$y = X\beta + \varepsilon. \quad (1)$$

We call this model as the Basic OLS model. Here, $y = (y_1, y_2, \dots, y_n)'$, where $y_i = (y_{i1}, y_{i2}, \dots, y_{i10})'$ is a vector of utility values of 10 health states elicited from the ith participant $(i = 1, 2, \dots, n)$, and n is the number of participants; $\beta = (\beta_0, \beta_1, \dots, \beta_{20})'$ is a vector of regression coefficients to be estimated, corresponding to the intercept and 20 indicator variables representing the four severity levels (2 to 5 levels) of the five dimensions (dummy coding scheme); X is an $10n \times 21$ design matrix with the first column being an identity vector for the intercept and the remaining 20 columns for the indicator variables; ε is an $10n \times 1$ vector of errors. It is assumed that ε_{ij} (error for utility value of jth health state in ith participant) ~ i.i.d. $N(0, \sigma^2)$, where i.i.d. stands for independent and identically distributed and $N(0, \sigma^2)$ denotes a normal distribution with mean 0 and variance σ^2.

As each participant values 10 health states, values of health states elicited by the same participant might be correlated. Furthermore, the variance may not be constant. Thus, it is advisable to use cluster-robust SE (considering each participant

as a cluster) to make valid statistical inference about the coefficients of the model.(Williams, 2000)

Alternatively, one can use an RE model for utility values specified as:(Verbeke, 2000)

$$y = X\beta + \gamma + \varepsilon. \quad (2)$$

We call this model as the Basic RE model. Here, y, X, and β are vectors as mentioned in equation (1); $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n)'$, where $\gamma_i = (\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{i10})'$, $\gamma_{ij} = \gamma_{ik}$ for all $j \neq k$, represents a 10×1 vector of ith participant-specific random intercept which underlines the intra-participant correlation among the 10 health states; ε is an $10n \times 1$ vector of errors. It is assumed that $\gamma_{ij} \sim N(0, \sigma_\gamma^2)$, $\varepsilon_{ij} \sim N(0, \sigma_e^2)$, and γ_i's and ε_{ij}'s are independent. Here, σ_γ^2 represents the between-participant variance, σ_e^2 is the within-participant variance, $Cov(y_{ij}, y_{ik}) = \sigma_\gamma^2$ for all $j \neq k$, $Cov(y_{ij}, y_{lk}) = 0$ for all $i \neq l$ and $Var(y_{ij}) = \sigma_\gamma^2 + \sigma_e^2 = \sigma^2$. In other words, within-participant utility values are correlated, but between-participant utility values are uncorrelated. Furthermore, intra-participant correlation does not depend on order of health states being valued.

The models in equations (1) and (2) may be extended to include additional variables, to indicate whether a health state is mild (e.g., all dimensions at severity level 1 or 2) or severe (e.g., at least one dimension at severity level 4 or 5), and possibly adding interactions between the health state descriptor variables to improve model performance for the country-specific valuation data.(Szende et al., 2007) Furthermore, indicator variables for health state descriptor may be coded with backward difference coding scheme,(Chen et al., 2003) so that regression coefficients of each dimension represent estimated mean differences in the utility values compared to their previous level of severity. In this thesis, the methodologies are presented for the dummy coding scheme for illustration as it is widely used. However, relevant information regarding the differential coding scheme is provided in **Publication Paper IV**.

4.5.3.4.2 Prerequisite Parameters for Sample Size Estimation

Inverse of cross-product of design matrix

As will be shown in the subsequent sections, the inverse of the cross-product of the design matrix $C = (X'X)^{-1}$, where X is as defined in equation (1), plays an important role in the sample size determination. Thus, this piece of generic information was derived before discussing specific sample size determination methods.

It is intuitive to begin with considering the sample size as a multiple of 10. Since the valuation protocol has only 10 unique blocks of health states (**Appendix 6**), each block has 10 health states, and the blocks have equal probability of being selected, for $N_s = 10 \times M$ participants, each of the 10 blocks on average is to be valued M times. Thus, for N_s participants, the matrix X is expected to be $X = \left(X_{u(1)}, X_{u(2)}, \ldots, X_{u(M)}\right)'$, where $X_{u(i)} = X_u$ is the *i*th replicate of the 100×21 design matrix based on the 10 unique blocks. Therefore,

$$E(C|N_s = 10M) = (X_u'X_u)^{-1}/M = D/M = 10\,D/N_s, \qquad (3)$$

where $D = (X_u'X_u)^{-1}$.

As such,

$$E\left(C_{jj}\right) = \frac{D_{jj}}{M} = \frac{10D_{jj}}{N_s}, \qquad (4)$$

where C_{jj} and D_{jj} are the *j*th diagonal element of the 21×21 matrices C and D, respectively.

The matrix D for the dummy coding scheme is shown in **Table 2**.

Table 2. *D* matrix based on indicator variables defined using the dummy coding scheme for EQ-5D-5L health state descriptors

	INT	MO2	MO3	MO4	MO5	SC2	SC3	SC4	SC5	UA2	UA3	UA4	UA5	PD2	PD3	PD4	PD5	AD2	AD3	AD4	AD5
INT	0.131																				
MO2	-0.037	0.117																			
MO3	-0.022	0.052	0.129																		
MO4	-0.029	0.057	0.058	0.155																	
MO5	-0.013	0.056	0.059	0.066	0.130																
SC2	-0.051	0.022	0.007	0.000	-0.008	0.111															
SC3	-0.011	-0.016	-0.011	-0.052	-0.033	0.045	0.148														
SC4	-0.020	-0.012	-0.024	-0.034	-0.021	0.051	0.067	0.140													
SC5	-0.019	-0.013	-0.025	-0.038	-0.042	0.052	0.068	0.074	0.124												
UA2	-0.050	0.011	-0.008	0.023	-0.018	0.015	-0.031	-0.001	0.001	0.125											
UA3	-0.041	-0.012	-0.014	-0.001	-0.013	-0.005	-0.018	-0.004	-0.001	0.066	0.136										
UA4	-0.039	-0.018	-0.011	-0.002	-0.036	0.006	0.005	0.013	0.019	0.063	0.061	0.141									
UA5	-0.028	-0.001	-0.010	0.001	-0.012	0.002	-0.014	0.000	-0.006	0.053	0.052	0.056	0.123								
PD2	-0.051	0.011	0.006	0.006	0.004	0.010	-0.004	-0.006	-0.012	-0.002	0.010	-0.006	0.007	0.101							
PD3	-0.007	-0.020	-0.008	-0.020	-0.022	-0.015	0.017	-0.015	-0.023	-0.037	-0.035	-0.017	0.004	0.036	0.150						
PD4	0.000	-0.019	-0.008	-0.008	-0.010	-0.013	-0.007	-0.029	-0.018	-0.024	-0.018	-0.018	0.005	0.034	0.059	0.114					
PD5	-0.035	-0.002	0.003	0.012	-0.034	0.006	-0.016	-0.029	-0.025	0.022	0.021	0.009	-0.019	0.041	0.037	0.040	0.138				
AD2	-0.063	-0.008	-0.006	0.007	-0.001	0.007	-0.028	-0.012	-0.009	0.019	0.015	0.016	0.005	0.009	-0.006	-0.015	0.008	0.126			
AD3	-0.050	0.006	-0.008	-0.010	0.013	-0.013	-0.031	-0.033	-0.028	-0.001	0.011	-0.015	-0.019	0.022	-0.011	-0.021	-0.001	0.068	0.156		
AD4	-0.045	0.000	-0.010	0.000	0.010	-0.010	-0.020	-0.007	-0.006	-0.007	-0.009	-0.022	-0.040	0.009	-0.011	-0.024	0.007	0.063	0.079	0.135	
AD5	-0.037	-0.009	-0.018	-0.027	-0.020	-0.003	0.003	-0.008	-0.009	-0.010	-0.016	-0.012	-0.034	0.006	0.007	-0.013	-0.012	0.059	0.077	0.071	0.124

EQ-5D-5L: 5-Level EuroQol 5-Dimension; INT: Intercept; MO2 to MO5, SC2 to SC5, UA2 to UA5, PD2 to PD5, and AD2 to AD5 represent elements corresponding to indicator variables for self-care, usual activity, pain/discomfort, and anxiety/depression dimension at severity level 2 to 4 (reference: severity level 1), respectively.

Design effect

Standard sample size estimation approaches, as well as those described in the subsequent sections, assume (or begin with assuming) independent observations. However, the EQ-VT protocol requires each participant to value 10 health states. Such clustering of observations within participants impacts the variance of the estimates. On the one hand, utility values (outcome variable) elicited by the same participant are likely positively correlated, as some people may generally give higher (or lower) values to health states no matter what these health states are. On the other hand, the EQ-VT protocol is developed to ensure that the 10 health states (exposure variables) are diverse within each participant, leading to negative correlation in the exposure. Given positive correlation in outcomes, the variances of the regression estimates are inflated if the exposures are positively correlated (e.g., cluster randomized trials), or deflated if the exposures are negatively correlated (e.g., crossover drug trials).(Kahan and Morris, 2013, Parzen et al., 1998) The sample size should be increased (or decreased) by the 'design effect' factor (DE)(Kahan and Morris, 2013, Donner and Klar, 2000, Parzen et al., 1998) to compensate for the variance inflation or deflation. The sample size adjusted for the design effect is calculated as,

$$N = N_s \times DE\ . \tag{5}$$

For a randomized controlled trial that has one exposure variable, there are formulas for estimation of design effect and therefore sample size using the intra-class correlation coefficients as inputs.(Kahan and Morris, 2013, Donner and Klar, 2000, Parzen et al., 1998) However, currently, there is no such formula for the complex case of multivariable regression analysis. Therefore, it was proposed to empirically estimate the design effect using data from a large-scale study, as will be shown below.

The design effect for predicted utility value of a particular health state HS_0 is calculated as

$$DE_{y_0} = \frac{V_{Cluster}(\hat{y}_0)}{V(\hat{y}_0)}, \tag{6}$$

where $V_{Cluster}(\hat{y}_0)$ and $V(\hat{y}_0)$ are cluster-robust and OLS variances, respectively, for the predicted utility value of HS_0. For simplicity, one may consider using simple or weighted mean of DE_{y_0} values ($\overline{DE}_y$) of directly valued health states, where weights can be chosen proportional to probability of occurrence of the health states. According to the valuation protocol, probabilities of occurrences of very mild health states (21111, 12111, 11211, 11121, and 11112), the most severe health state (55555) and the remaining 80 directly valued health states are 0.02, 1.0, and 0.01, respectively (**Appendix 6**).

Similar to equation (6), the design effect for a regression coefficient, β_j, is calculated as

$$DE_{\beta_j} = \frac{V_{Cluster}(\hat{\beta}_j)}{V(\hat{\beta}_j)}, \quad (7)$$

where $V_{Cluster}(\hat{\beta}_j)$ and $V(\hat{\beta}_j)$ are cluster-robust and OLS variances of β_j, respectively. Similar to $\overline{DE}_y$, one may consider using simple or weighted mean of DE_{β_j}values ($\overline{DE}_\beta$). For example, if we are interested in coefficients of health state descriptors (i.e., β_1 to β_{20}), weights proportional to probability of occurrence of respective severity levels in a sample of randomly allocated blocks can be used for $\overline{DE}_\beta$. According to the valuation protocol, the probability of occurrences of severity levels 2 to 5 in different dimensions range from 0.13 to 0.26 (**Appendix 6**).

As an alternative to $V_{Cluster}(\hat{y}_0)$ and $V_{Cluster}(\hat{\beta}_j)$ in equations (6) and (7) respectively, one can use variances estimated for predicted utility value of HS_0, $V_{RE}(\hat{y}_0)$, and β_j, $V_{RE}(\hat{\beta}_j)$, respectively, from the Basic RE model in equation (2). Empirical estimates of design effects from the study C will be given in Results section.

4.5.3.4.3 Sample Size Estimation Approaches for Value Set Studies

Approach 1 - To achieve desired precision for an estimated mean utility value for a particular health state

The primary objective of value set studies is to predict utilities for health states with acceptable precision. Suppose the study team would like to predict utility value of a particular health state HS_0 with a tolerated margin of error δ, where δ is a sufficiently

small value (e.g., MID for utility values).(Coretti et al., 2014) That is, the maximum allowable difference between the predicted value and the true value should be less than δ with sufficiently high probability $100 \times (1-\alpha)\%,\ 0 < \alpha < 1$. This is equivalent to targeting the $100 \times (1-\alpha)\%$ prediction interval (PI) to be $\pm\,\delta$.

Consider the Basic OLS model for estimating the mean utility value of a particular health state, *HS₀*. The $100 \times (1-\alpha)\%$ PI for mean utility value of *HS₀* is:(Montgomery et al., 2012)

$$\left(\hat{y}_0 - Z_{1-\frac{\alpha}{2}}\sqrt{\hat{\sigma}^2\, x_0' C x_0},\ \hat{y}_0 + Z_{1-\frac{\alpha}{2}}\sqrt{\hat{\sigma}^2\, x_0' C x_0}\right), \tag{8}$$

where $\hat{y}_0$ is the predicted utility value for *HS₀*, $x_0' = (1, x_{01}, x_{02}, \ldots, x_{020})$ is the vector representing values of indicator variables for health state descriptor of *HS₀*, $\hat{\sigma}^2$ is the estimate of error variance, and $C = (X'X)^{-1}$.

Thus, for the objective of estimating mean utility value of *HS₀* with desired precision, say $100 \times (1-\alpha)\%$ PI of $\hat{y}_0$ equal to $\hat{y}_0 \pm \delta$, the sample size N_s can be calculated by solving equation, $\delta = Z_{1-\frac{\alpha}{2}}\sqrt{\hat{\sigma}^2\, x_0' C x_0}$. It can be solved using equation (3) as $\delta = Z_{1-\frac{\alpha}{2}} \times \hat{\sigma} \times \sqrt{10\, x_0' D x_0 / N_s}$ giving:

$$N_s = \frac{10\,\hat{\sigma}^2\, x_0' D x_0\ Z^2_{1-\alpha/2}}{\delta^2}. \tag{9}$$

This approach provides different sample sizes for different health states. The study team may choose the mean of these sample sizes. Using the matrix D (**Table 2**) and x_0 vectors of the directly valued health states in the EQ-VT protocol, it can be shown that $x_0' D x_0$ has mean, standard deviation (SD) and coefficient of variation (CV) of 0.210, 0.062, and 29.3%, respectively.

Lastly, as equation (9) also assumes independent observations, it needs to be adjusted for the design effect $\overline{DE}_y$ using equation (5), and replacing $x_0' D x_0$ by mean of $x_0' D x_0$ values $(\bar{D}_{x_0})$ of directly values health states, gives

$$N = N_s \times DE = \frac{10\,\hat{\sigma}^2\, \bar{D}_{x_0}\, \overline{DE}_y\ Z^2_{1-\alpha/2}}{\delta^2}. \tag{10}$$

Approach 2 - Empirical approach to achieve desired mean absolute error in prediction of utility values relative to a reference study

This is an empirical approach in which mean absolute error (MAE)(Lamers et al., 2006) defined as,

$$MAE_{N_s} = \frac{1}{10N}\sum_{i=1}^{N}\sum_{j=1}^{10}\left|\hat{y}_{ij(N_s)} - y_{ij}\right|, \quad (11)$$

is calculated, where $\hat{y}_{ij(N_s)}$ denote predicted utility values based on a regression model fitted with sample size N_s, and y_{ij} denote observed utility values in a reference study with a larger sample size N, where $N > N_s$. $MAE_{N_s}$is to be estimated for a large number of replications of random sampling (with replacement) of N_s number of subjects within a reference study with sample size N conducted according to the EQ-VT protocol. A plot of mean MAE_{N_s}over sample size N_s can provide a visual presentation of how the MAE is reduced with an increase in sample size for the model. A sample size N_s^* corresponding to desired MAE, or desired marginal gain in MAE as sample size increases, is selected for the study.

Approach 3 - To assess significance of a regression coefficient of health state descriptors

A value set study is expected to show that individual regression coefficients of health state descriptors (i.e., β_1 to β_{20}) are not different from zero by chance. That is, it requires to test the null hypothesis $H_0: \beta_j = 0$ against the alternative hypothesis $H_1: \beta_j \neq 0$, for all $j = 1, 2\ldots, 20$, with sufficient statistical power $(1-\beta)$ and type-I error rate (α).

Considering the Basic OLS model in equation (1), the SE of $\hat{\beta}_j$ is $SE(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 C_{jj}}$, where $\hat{\sigma}^2$ is the estimate of error variance, and C_{jj} is the *j*th diagonal element of $C = (X'X)^{-1}$ matrix [15]. Using equation (4), $SE(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 D_{jj}/M} = \sqrt{10\hat{\sigma}^2 D_{jj}/N_s}$, where D_{jj} are available from Table 2. Solving the equation, $\beta_j/SE(\hat{\beta}_j) = \beta_j/\sqrt{10\hat{\sigma}^2 D_{jj}/N_s} = (Z_{1-\alpha/2} + Z_{1-\beta})$ will give sample size:

$$N_s = \frac{10\hat{\sigma}^2 D_{jj}(Z_{1-\alpha/2}+Z_{1-\beta})^2}{\beta_j^2}, \quad (12)$$

where $Z_{1-\alpha/2}$ and $Z_{1-\beta}$ are values of standard normal distribution at $1-\alpha/2$ and $1-\beta$, respectively. The $\hat{\sigma}^2$ can be obtained from a reference study (e.g., study C).

The target β_j in sample size determination can be a smallest desired value for the coefficient (e.g., MID for utility values) that the study team would consider meaningful. It is clear that the sample size may be different for different coefficients; one can choose the mean of coefficients $(\bar{\beta})$for simplicity in the sample size estimation. Substituting $\bar{\beta}$ for β_j and $\bar{D}$ for D_{jj} in equation (12) gives,

$$N_s = \frac{10\,\hat{\sigma}^2\,\bar{D}\,(Z_{1-\alpha/2}+Z_{1-\beta})^2}{\bar{\beta}^2}. \tag{13}$$

Since the sample size formula assumes independent observations, it needs to be adjusted for the design effect $\overline{DE}_\beta$ using equations (5), which gives,

$$N = \frac{10\,\hat{\sigma}^2\,\bar{D}\,\overline{DE}_\beta\,(Z_{1-\alpha/2}+Z_{1-\beta})^2}{\bar{\beta}^2}. \tag{14}$$

Approach 4 - To estimate a regression coefficient of health state descriptors with a desired precision

Alternative to Approach 3, the study team may desire to estimates the regression coefficients with certain precision, say requiring the $100\times(1-\alpha)\%$ CI of β_j to be $\hat{\beta}_j \pm \delta_j$, where δ_j is a sufficiently small number (e.g., MID for utility values). For this objective, the sample size is obtained by solving $\delta_j = Z_{1-\frac{\alpha}{2}}SE(\hat{\beta}_j) = Z_{1-\frac{\alpha}{2}}\sqrt{10\hat{\sigma}^2 D_{jj}/N_s}$ (Montgomery et al., 2012). It gives

$$N_s = \frac{10\,\hat{\sigma}^2 D_{jj} Z_{1-\alpha/2}^2}{\delta_j^2}. \tag{15}$$

Alternatively, substituting δ_j by $\bar{\delta}$ = mean of δ_j $(j = 1, 2, .., 20)$, D_{jj} by $\bar{D}$ in equation (15) gives

$$N_s = \frac{10\;\hat{\sigma}^2\,\bar{D} Z_{1-\alpha/2}^2}{\bar{\delta}^2}. \tag{16}$$

Finally, adjusting the sample size for the design effect $\overline{DE}_\beta$ using equation (5), gives

$$N = \frac{10\,\hat{\sigma}^2\,\bar{D}\,\overline{DE}_\beta\,Z_{1-\alpha/2}^2}{\bar{\delta}^2}. \tag{17}$$

5 RESULTS

5.1 Results for Research Aim 1

In the sample for the SF-6D value set in study A, 7 participants valued dead higher than all the other states; 1 participant valued the 'all-worst' state higher than all the other states, and 5 participants were observed to have a poor understanding of health states description and/or valuation tasks. Hence, these 13 participants were excluded from the SF-6D-related analysis. **Table 3** shows sociodemographic and health characteristics of 1021 participants for the SF-6D valuation that were included in the analysis. Due to the pre-specified quota for gender, age and ethnicity, the demographic characteristics of enrolled participants were similar to what was planned. The majority of the SF-6D participants were married (n = 765, 75%), employed/self-employed (n = 659, 65%), had at least secondary education (n = 844, 83%), and self-reported good to excellent general health (n = 947, 93%).

Similarly, in the EQ-5D-3L sample, 12 participants valued dead higher than all the other states; 2 participants valued the unconscious state higher than all the other states; 1 participant did not value the 'all-worst' state, and 4 participants were observed to have a poor understanding of health states description and/or valuation tasks. Hence, 19 participants were excluded from the EQ-5D-3L-related analysis. The sociodemographic and health characteristics of the EQ-5D-3L participants were similar to the SF-6D participants (**Table 3).**

Table 3. Sociodemographic and health characteristics of study A participants

Characteristics	SF-6D sample (N = 1021)	EQ-5D-3L sample (N = 1015)
Female, n (%)	521 (51.0)	512 (50.4)
Age (years), n (%)		
21-59	890 (87.2)	891 (87.8)
60+	131 (12.8)	124 (12.2)
Ethnicity, n (%)		
Chinese	392 (38.4)	387 (38.1)
Malay	396 (38.8)	399 (39.3)
Indian	233 (22.8)	229 (22.6)
Education level, n (%)		
Primary (6 years) or less	177 (17.3)	190 (18.7)
Secondary (11 years)	562 (55.0)	576 (56.8)
Diploma/degree or higher	282 (27.6)	249 (24.5)
Married/partner, n (%)	765 (74.9)	761 (75.0)
Employed or self-employed, n (%)	659 (64.5)	643 (63.4)
Self-reported health on VAS, Mean (SD)	84.5 (11.2)	83.0 (12.1)

VAS Visual analog scale (100: Best imaginable health state, 0: Worst imaginable health state); SD Standard deviation.

In the SF-6D valuation, except the 'all-worst' state, no other health state was valued worse than dead state. The mean valuation scores for selected SF-6D health states are shown in **Figure 4**. For the participants who considered the 'all-worst' state worse than dead, there was a difference of more than 30 points in the mean valuation score between the 'all-worst' state ('645655') and its adjacent health states, which are only one level different in one dimension ('545655' and '645555'). For the participants who considered the 'all-worst' state better than dead, the corresponding difference ranged from 9 to 17 points. Similar to the SF-6D results, only the 'all-worst' state was valued worse than dead, and the majority of participants (n = 753, 74%) considered the 'all-worst' state worse than dead in the EQ-5D-3L valuation. Similar to **Figure 4**, **Figure 5** also shows a difference of 25 points in the mean valuation score between the 'all-worst' state ('33333') and its adjacent state '33323' for the participants who considered the 'all-worst' worse than dead. For the participants who considered the 'all-worst' better than dead the corresponding difference was 11 points.

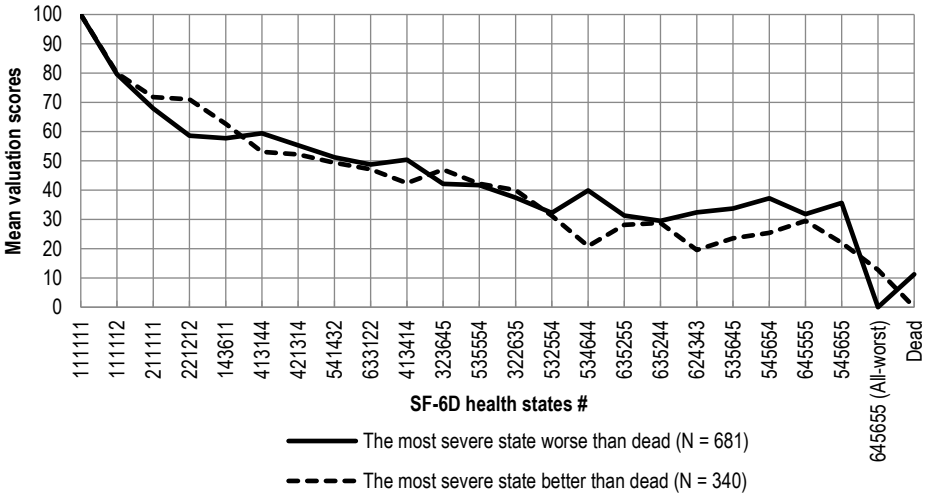


Figure 4. Mean valuation scores of SF-6D health states elicited by participants who considered the 'All-worst' state better and worse than dead

Some health states were systematically skipped for the graphical presentation, except the lowest ten health states with the least valuation scores near the dead state.

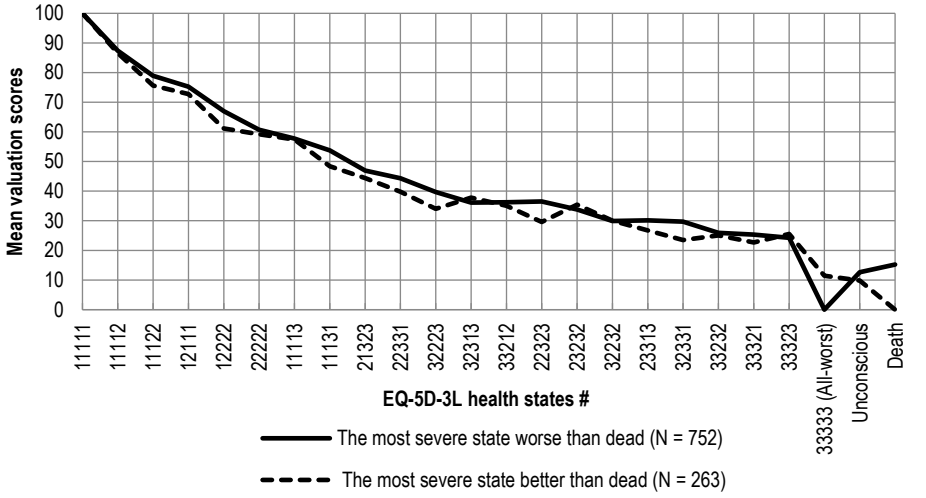


Figure 5. Mean valuation scores of EQ-5D-3L health states elicited by participants who considered the 'All-worst' state better and worse than dead

Some health states were systematically skipped for the graphical presentation, except the lowest ten health states with the least valuation scores near the dead state.

In the OLS model for SF-6D valuation score (**Table 4**), the coefficient D1 showed that the participants who considered the 'all-worst' state worse than death, scored higher in the other health states by 2 points (95% CI: 0.2, 3.8); compared to the participants who considered the 'all-worst' state better than dead. The participants who considered the 'all-worst' state better than dead scored the 'all-worst' state 12.3 points lower (D2 coefficient 95% CI: −15.8, −8.8; p <0.001) than expected by its descriptors. Furthermore, the participants who considered the 'all-worst' state worse than dead scored the 'all-worst' state 27.0 points lower (coefficient: −12.3 - 14.7 = 27.0; 95% CI: −30.2, −24.0; p <0.001) than expected.

For the EQ-5D-3L valuation, the participants who considered the 'all-worst' state better than dead did not score the 'all-worst' state statistically significantly different from what is expected by its descriptors (coefficient = -2.1 points; 95% CI: −5.4, −1.1; p = 0.201). The participants who considered the 'all-worst' state worse than dead scored the 'all-worst' state 15.8 points lower (coefficient = −2.1-13.7 = −15.8; 95% CI: −18.8, −12.7; p <0.001) than expected (**Table 5**).

Table 4. Summary of ordinary least-square regression model for SF-6D valuation score

Regressor	Coefficient (95% Confidence interval)
The most severe state worse than dead (D_1)	2.0 (0.2, 3.8) *
The most severe state (D_2)	-12.3 (-15.8, -8.8) ***
Interaction of D_1 and D_2	-14.7 (-16.6, -12.9) ***
At least one severe level (N_3)	-2.3 (-4.3, -0.4) *
Physical functioning level 2	-7.5 (-16.6, -12.9) ***
Physical functioning level 3	-7.9 (-9.7, -6.0) ***
Physical functioning level 4	-13.8 (-15.8, -11.7) ***
Physical functioning level 5	-13.9 (-15.9, -11.9) ***
Physical functioning level 6	-22.1 (-24.4, -19.9) ***
Role limitations level 2	-3.7 (-5.2, -2.2) ***
Role limitations level 3	-3.1 (-4.6, -1.6) ***
Role limitations level 4	-2.7 (-4.3, -1.0) **
Social functioning level 2	-3.1 (-4.6, -1.5) ***
Social functioning level 3	-2.1 (-3.8, -0.4) *
Social functioning level 4	-2.9 (-4.8, -1.1) **
Social functioning level 5	-3.2 (-5.1, -1.4) **
Pain level 2	-3.9 (-5.7, -2.2) ***
Pain level 3	-6.3 (-8.1, -4.5) ***
Pain level 4	-6.8 (-8.6, -4.9) ***
Pain level 5	-7.5 (-9.2, -5.7) ***
Pain level 6	-11.1 (-13.0, -9.2) ***
Mental health level 2	-2.1 (-3.8, -0.4) *
Mental health level 3	-2.8 (-4.7, -1.0) **
Mental health level 4	-4.0 (-5.9, -2.1) ***
Mental health level 5	-3.0 (-4.9, -1.1) **
Vitality level 2	-0.3 (-2.1, 1.6)
Vitality level 3	-5.0 (-6.9, -3.2) ***
Vitality level 4	-5.6 (-7.6, -3.6) ***
Vitality level 5	-10.6 (-12.6, -8.6) ***
Intercept	80.2 (77.3, 83.1) ***
$R^2 = 0.519$	

* $p < 0.05$, ** $p < 0.01$, *** $p<0.001$.

Table 5. Summary of ordinary least-square regression model for EQ-5D-3L valuation score

Regressor	Coefficient (95% Confidence interval)
The most severe state worse than dead (D_1)	2.2 (-0.0, 4.5)
The most severe state (D_2)	-2.1 (-5.4, 1.1)
Interaction of D_1 and D_2	-13.7 (-15.8, -11.6) ***
At least one severe level (N_3)	-14.8 (-17.4, -12.1) ***
Mobility level 2	-8.2 (-9.4, -6.9) ***
Mobility level 3	-15.0 (-17.0, -13.0) ***
Self-care level 2	-5.1 (-6.5, -3.7) ***
Self-care level 3	-13.9 (-15.7, -12.2) ***
Usual activities level 2	-1.6 (-3.5, 0.3)
Usual activities level 3	-4.2 (-6.3, -2.0) ***
Pain/discomfort level 2	-5.1 (-6.3, -3.9) ***
Pain/discomfort level 3	-12.4 (-13.6, -11.1) ***
Anxiety/depression level 2	-0.3 (-1.8, 1.2)
Anxiety/depression level 3	-6.1 (-7.9, -4.2) ***
Intercept	79.9 (77.5, 82.3) ***
$R^2 = 0.642$	

* $p < 0.05$, ** $p < 0.01$, *** $p<0.001$.

5.2 Results for Research Aim 2

In the sample for the EQ-5D value set in study A, 9 participants had chronic diseases other than diabetes, high blood pressure/hypertension, heart disease, asthma/lung disease or rheumatism/back pain/other bone-muscle illness; 34 participants valued 'dead' state higher than all the other states; 3 participants valued 'unconscious' state higher than all the other states; 1 participant did not value the 'all-worst' state, and 4 participants were observed to have a poor understanding of valuation tasks. Hence, a total of 51 participants were excluded from the analysis (i.e., included 1034 – 51 = 983 participants). **Table 6** shows the comparison of sociodemographic and health characteristics of the patient groups with the control group. Participants in the patient's groups were older, had a lower education level, and lower self-reported general health, compared to participants in the control group.

Table 6. Comparison of sociodemographic and health characteristics of patient groups with the control group in the EQ-5D-3L sample of study A

Characteristics	Control (N = 651)	Diabetes (N = 102)		Rheumatism (N = 162)		Hypertension (N = 145)		Heart diseases (N = 44)		Lung diseases (N = 44)	
	n (%)	n (%)	*P*	n (%)	*P*	n (%)	*p*	n (%)	*p*	n (%)	*p*
Female	329 (50.5)	47 (46.1)	0.404	92 (56.8)	0.071	69 (47.6)	0.530	9 (20.5)	<0.001	29 (65.9)	0.044
Age (years)			<0.001		<0.001		<0.001		<0.001		0.048
21-29	168 (25.8)	0 (0.0)		7 (4.3)		2 (1.4)		0 (0.0)		14 (31.8)	
30-39	178 (27.3)	6 (5.9)		21 (13.0)		7 (4.8)		1 (2.3)		9 (20.5)	
40-49	181 (27.8)	22 (21.6)		43 (26.5)		24 (16.6)		7 (15.9)		5 (11.4)	
50-59	91 (14.0)	41 (40.2)		40 (24.7)		57 (39.3)		18 (40.9)		8 (18.2)	
60+	33 (5.1)	33 (32.4)		51 (31.5)		55 (37.9)		18 (40.9)		8 (18.2)	
Ethnicity			0.001		0.024		0.914		0.001		0.287
Chinese	234 (35.9)	26 (25.5)		67 (41.4)		56 (38.6)		9 (20.5)		16 (36.4)	
Malay	284 (43.6)	37 (36.3)		50 (30.9)		57 (39.3)		14 (31.8)		14 (31.8)	
Indian	133 (20.4)	39 (38.2)		45 (27.8)		32 (22.1)		21 (47.7)		14 (31.8)	
Education level			<0.001		<0.001		<0.001		<0.001		0.352
Primary (6 years) or less	74 (11.4)	48 (47.1)		60 (37.0)		58 (40.0)		22 (50.0)		12 (27.3)	
Secondary (11 years)	387 (59.5)	44 (43.1)		79 (48.8)		77 (53.1)		18 (40.9)		22 (50.0)	
Diploma/degree or higher	190 (29.2)	10 (9.8)		23 (14.2)		10 (6.9)		4 (9.1)		10 (22.7)	
Married/living with partner	481 (73.9)	84 (82.4)	0.090	128 (79.0)	0.233	120 (82.8)	0.022	38 (86.4)	0.106	23 (52.3)	0.001
Religion			0.051		0.094		0.796		0.017		0.894
No religion	52 (8.0)	4 (3.9)		12 (7.4)		11 (7.6)		3 (6.8)		3 (6.8)	
Buddhism/Taoism	139 (21.4)	19 (18.6)		43 (26.5)		37 (25.5)		4 (9.1)		12 (27.3)	
Islam	289 (44.4)	41 (40.2)		53 (32.7)		63 (43.5)		17 (38.6)		16 (36.4)	
Hinduism/Sikhism	117 (18.0)	31 (30.4)		40 (24.7)		24 (16.6)		17 (38.6)		10 (22.7)	
Christianity	54 (8.3)	7 (6.9)		14 (8.6)		10 (6.9)		3 (6.8)		3 (6.8)	
House type			0.377		0.175		0.786		0.204		0.533
Government owned: 4 rooms or smaller	449 (69.0)	71 (69.6)		106 (65.4)		96 (66.2)		25 (56.8)		27 (61.4)	
Government owned: 5 rooms or bigger	191 (29.3)	27 (26.5)		49 (30.3)		45 (31.0)		18 (40.9)		16 (36.4)	
Private	11 (1.7)	4 (3.9)		7 (4.3)		4 (2.8)		1 (2.3)		1 (2.3)	
Self-reported health on VAS, Mean (SD)	85.7 (10.4)	75.5 (14.4)	<0.001	75.5 (13.9)	<0.001	75.8 (14.6)	<0.001	72.3 (17.2)	<0.001	75.3 (16.1)	<0.001

VAS Visual analog scale (100: Best imaginable health state, 0: Worst imaginable health state); SD Standard deviation.

Of 525 participants who completed the interview in study B, 30 were excluded from analyses: 24 participants assigned the same value to all 10 health states, and six participants valued all the health states worse than or equal to death. **Table 7** shows the sociodemographic and health characteristics of the 495 participants included in the analyses (169 in the control group, 157 in the heart disease patients group, and 169 in the cancer patients group). Participants in the patient groups were older, had lower education level, and poor self-reported general health, compared to participants in the control group. A total of 62 (37%) participants in the control group self-reported to have one or more chronic diseases, such as hypertension (15%), hyper/dyslipidemia (12%), and lung diseases (10%). The majority of sociodemographic characteristics of the control group were similar to the census population (DOS, 2010). Only 5% of patients in the heart disease group self-reported to have had cancer. Similarly, 4% of patients in the cancer group had had heart disease, and less than 5% of participants in the control group had either heart disease or cancer. More details on heart disease and cancer patients' disease characteristics are presented in the **Online Appendix Tables 3** and **4** of **Publication Paper III**, respectively.

Table 7. Sociodemographic and health characteristics of study B participants

Characteristic	Control (N = 169)	Heart disease (N = 157)		Cancer (N = 169)	
	n (%)	n (%)	*p*	n (%)	*p*
Female	89 (52.7)	54 (34.4)	0.001	98 (58.0)	0.381
Age (years)			<0.001		<0.001
21-40	75 (44.4)	13 (8.3)		20 (11.8)	
41-60	65 (38.5)	61 (38.9)		93 (55.0)	
>60	29 (17.2)	83 (52.9)		56 (33.1)	
Ethnicity			0.029		0.079
Chinese	140 (82.8)	110 (70.1)		123 (72.8)	
Malay	10 (5.9)	12 (7.6)		22(13.0)	
Indian	10 (5.9)	23 (14.7)		10 (5.9)	
Others	9 (5.3)	12 (7.6)		14 (8.3)	
Education level			<0.001		<0.001
Primary (6 years) or less	12 (7.1)	35 (22.3)		25 (14.8)	
Secondary (7-11 years)	54 (32.0)	88 (56.1)		91 (53.9)	
Diploma, University or higher	103 (61.0)	34 (21.7)		53 (31.4)	
Married/living with partner	104 (61.5)	101 (64.3)	0.647	113 (66.9)	0.364
Employed	109 (64.5)	84 (53.5)	0.055	95 (56.2)	0.148
Household earnings per month			<0.001		0.010
<S$4000	60 (35.5)	98 (62.4)		87 (51.5)	
≥S$4000	91 (53.9)	44 (28.0)		65 (38.5)	
Don't know/refused	18 (10.7)	15 (9.6)		17 (10.1)	
Religion			0.693		<0.001
No religious belief	37 (21.9)	26 (16.6)		15 (8.9)	
Buddhism/Taoism	55 (32.5)	54 (34.4)		66 (39.1)	
Islam	15 (8.9)	19 (12.1)		34 (20.1)	
Christians	44 (26.0)	43 (27.4)		46 (27.2)	
Others	18 (10.7)	15 (9.6)		8 (4.7)	
Self-reported health on VAS, Mean (SD)	82.6 (10.4)	73.8 (14.5)	<0.001	76.3 (17.3)	<0.001

VAS Visual analog scale (100: Best imaginable health state, 0: Worst imaginable health state); SD Standard deviation.

Table 8 summarizes the comparison of health state valuation scores between the patient groups and the control group in Study A. After taking health state descriptors and covariates into account in the regression models, the mean differences between the patient groups and the control group regarding valuation scores of all the health states ranged from −2.5 to 1.6 (each $p > 0.05$), except for the heart disease group. The adjusted mean valuation score of all the health states for the heart disease group was 4.6 points higher (95% CI: 0.4 to 8.9; $p = 0.032$) than that of the control group. Similarly, the mean differences between the patient groups and the control group regarding severe health state valuation scores ranged from −2.4 to 1.8 (each $p > 0.05$), except for the heart disease group. The adjusted mean valuation score of severe states for the heart disease group was 5.4 points higher (95% CI: 0.7 to 10.1; $p = 0.025$) than that of the control group. There was no practically significant difference in the mean valuation scores of non-severe health states between any patient group and the control group.

Table 8. Comparison of valuation score between patient groups and the control group in study A

EQ-5D-3L health states#	Diabetes (N = 102)	Rheumatism (N = 162)	Hypertension (N = 145)	Heart diseases (N = 44)	Lung diseases (N = 44)	Control (N = 651)
All health states						
Mean (SD)	43.6 (30.0)	43.4 (30.6)	44.4 (30.0)	43.1 (28.7)	40.6 (29.5)	43.9 (30.6)
Adjusted mean difference (95% CI)	1.6 (-1.2, 4.3)	0.4 (-2.0, 2.8)	0.7 (-1.7, 3.1)	4.6 (0.4, 8.9)*	-2.5 (-6.2, 1.2)	-
Non-severe health state						
Mean (SD)	71.6 (18.8)	71.3 (20.0)	71.2 (19.7)	69.8 (16.8)	69.0 (22.1)	72.7 (19.3)
Adjusted mean difference (95% CI)	1.0 (-2.5, 4.6)	-0.3 (-3.4, 2.9)	-1.0 (-4.3, 2.3)	2.6 (-2.4, 7.6)	-2.6 (-9.0, 3.8)	-
Severe health states						
Mean (SD)	31.4 (25.4)	31.9 (26.6)	33.6 (26.5)	33.7 (26.0)	28.7 (23.6)	31.5 (25.8)
Adjusted mean difference (95% CI)	1.8 (-1.2, 4.7)	0.7 (-1.9, 3.3)	1.3 (-1.2, 3.9)	5.4 (0.7, 10.1)*	-2.4 (-6.1, 1.2)	-

Health states with at least one domain at severity level 3 are classified as 'severe,' remaining health states are classified as 'non-severe.'

Adjusted mean difference: mean scores of the patient groups minus mean scores of the control group, estimated by the ordinary least-square regression model, with adjustment for health state descriptors, disutility due to severe problems, and sociodemographic covariates (see Methods section).

* $p < 0.05$.

SD Standard deviation; CI Confidence interval.

Table 9 summarizes the comparison of utility values between the patient groups and the control group in study B. After taking sociodemographic variables into account in the regression models, the mean utility value based on heart disease patients was lower by 0.05 (p = 0.534) for severe health states, and higher by 0.08 points (p = 0.007) and 0.09 points (p = 0.176) for mild and moderate health states, respectively, than that based on the control group. Unlike heart disease patients, there was no statistically significant difference in mean utility value for all health states including mild, moderate, and severe health states between cancer patient group and the control group.

Among all the covariates adjusted for the analysis, age had the strongest association with utility (Wald test p <0.0001; see **Online Appendix Table 5** of **Publication Paper III**). Older age was associated with lower utility value.

Table 9. Comparison of utility values between the patients and the control groups in study B

EQ-5D-5L health states#	Heart disease (N=157)	Cancer (N=169)	Control (N=169)
All health states			
Mean (SD)	0.291 (0.807)	0.336 (0.723)	0.398 (0.657)
Adjusted mean difference (95% CI)	0.031 (-0.067, 0.130)	-0.007 (-0.058, 0.045)	
Mild health states			
Mean (SD)	0.875 (0.345)	0.829 (0.362)	0.842 (0.255)
Adjusted mean difference (95% CI)	0.076 (0.021, 0.131) **	-0.009 (-0.072, 0.054)	
Moderate health states			
Mean (SD)	0.452 (0.704)	0.417 (0.651)	0.505 (0.545)
Adjusted mean difference (95% CI)	0.091 (-0.041, 0.223)	-0.011 (-0.132, 0.111)	
Severe health states			
Mean (SD)	-0.267 (0.764)	-0.096 (0.719)	-0.016 (0.693)
Adjusted mean difference (95% CI)	-0.048 (-0.198, 0.103)	-0.002 (-0.137, 0.133)	

Health states with all dimensions at severity level either 1 or 2 are considered 'mild,' health states with at least one dimension at severity level either 4 or 5 are considered 'severe,' and remaining are considered 'moderate.'

Adjusted mean difference: mean scores of the patient groups minus mean scores of the control group estimated by the ordinary least-square regression model, with adjustment for sociodemographic covariates (see Methods section).

* p <0.05; ** p <0.01.

SD Standard deviation; CI Confidence interval.

Mean (unadjusted and adjusted) differences between the heart disease patients and the general population group in differences between selected health states are presented in **Figure 6**. It shows that the covariate-adjusted differences in differences were higher than 0.05, a possible MID for utility values, for 23 of 45 pairs of health states (51%; p <0.05 for 11 pairs), and lower than 0.05 for 13 pairs of health states (29%; p <0.05 for 3 pairs) in heart disease patients than the control group. A similar analysis comparing cancer patients and the control group, (**Figure 7**) showed that adjusted differences in differences were greater than 0.05 for 7 of 45 pairs of health states (16%; p >0.05 for all), and lower than 0.05 for 4 pairs (9%; p >0.05 for all).

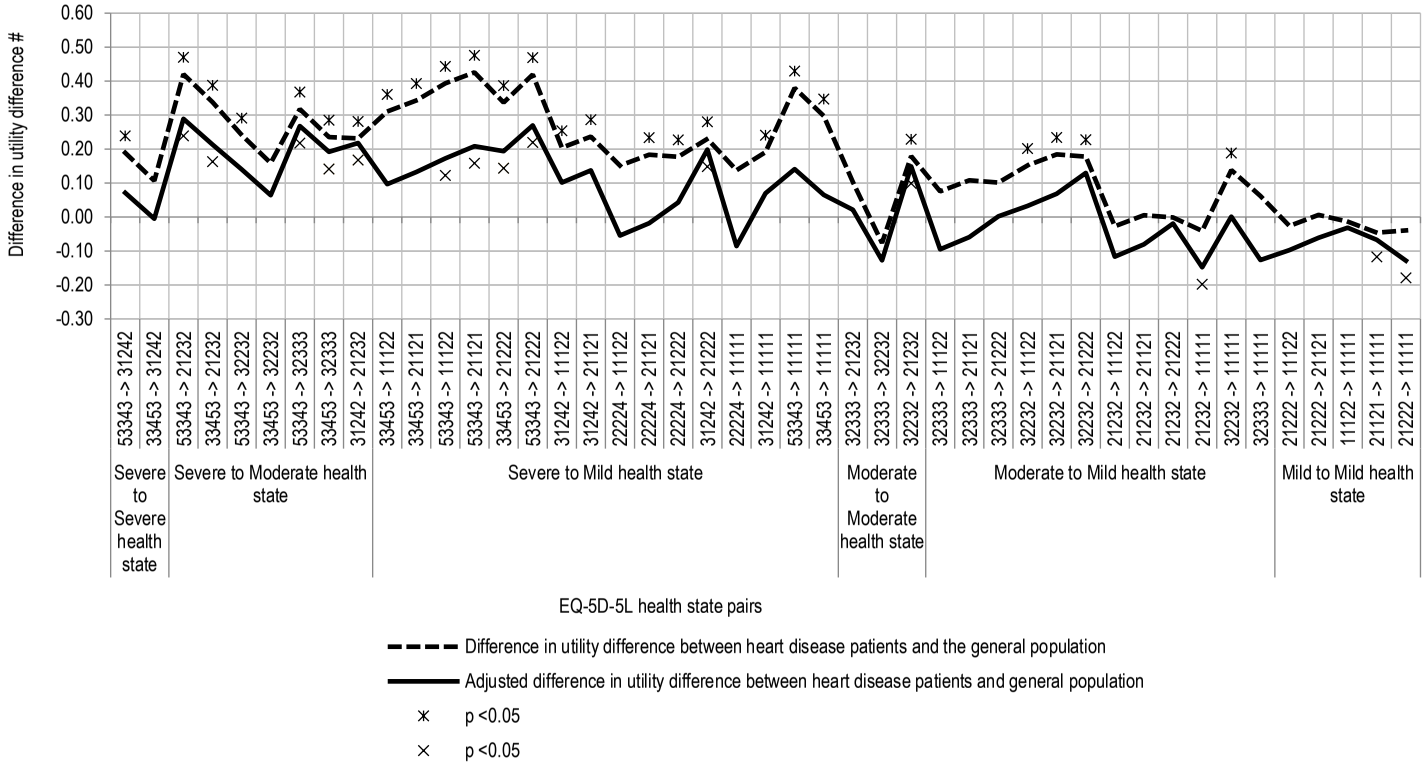


Figure 6. Estimated differences in differences in utility values between the heart disease and the control groups for pairs of health states

Differences in utility values were defined by subtracting the utility value of a poorer health state from the utility value of a better health state.

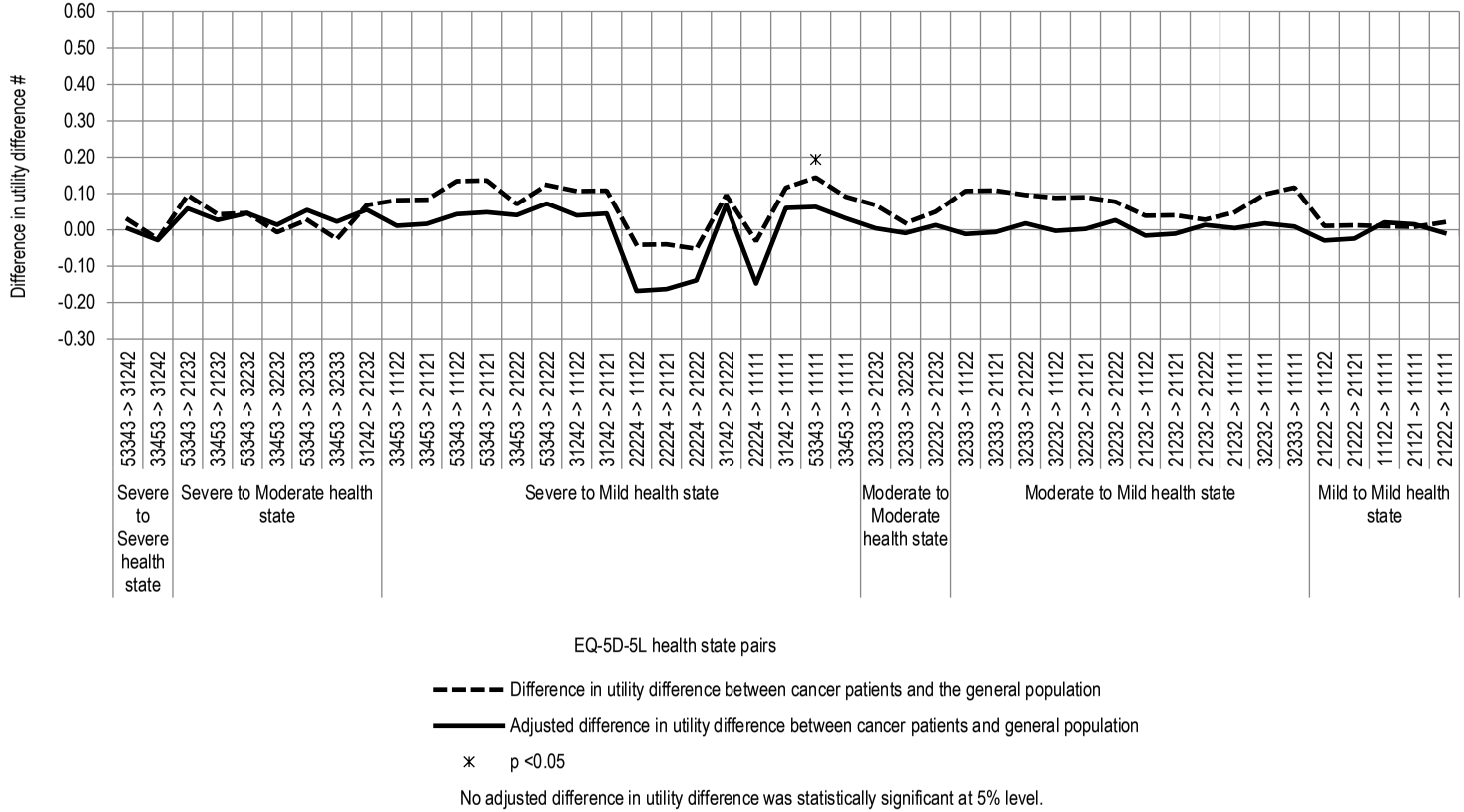


Figure 7. Estimated differences in differences in utility values between cancer and the control groups for pairs of health states

Differences in utility values were defined by subtracting the utility value of a poorer health state from the utility value of a better health state.

5.3 Results for Research Aim 3

5.3.1 Estimates of Useful Parameters

Based on the utility values from study C for the Basic OLS model in equation (1), the estimate of error variance ($\hat{\sigma}^2$) was 0.365. The mean ($\bar{\beta}$) of regression coefficients ($\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_{20}$) of health state descriptor defined using the dummy coding scheme was −0.196. Using the Basic RE model in equation (2), the estimate of between-participant variance ($\hat{\sigma}_\gamma^2$) and the within-participant variance ($\hat{\sigma}_e^2$) were 0.203 and 0.162, respectively. The Basic RE model gave $\overline{\beta}$ −0.204, which was similar to the -0.196 from the Basic OLS model.

Simple and weighted means of design effects ($\overline{DE}_\beta$) for $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_{20}$, using equation (7) for cluster-robust variances, were 0.82 and 0.84, respectively, or approximately 0.8. The Basic RE model was more efficient, in the sense that the variance estimates were smaller. Simple and weighted mean design effects based on RE variances were both 0.47, or approximately 0.5. Design effects for the intercept, $\hat{\beta}_0$, were 1.27 and 0.92 based on cluster-robust and RE variances, respectively. Detailed results of the Basic OLS and RE models are presented in **Electronic Supplementary Material Table A.2** of **Publication Paper IV**.

Pearson's correlation between predicted utility values based on the Basic OLS and RE models was 0.999 in study C, and mean absolute error in prediction using both the models was 0.499. Mean design effects ($\overline{DE}_{y_0}$) for predicted values using equation (6) were approximately 1.15 and 0.75 based on cluster-robust and RE variances respectively. Weighted means of design effects were similar to their unweighted mean design effects. Mean utility values of individual health states in study C along with their predicted values and their standard errors (square root of variances) using the Basic OLS and RE models are presented in **Electronic Supplementary Material Table A.4** of **Publication Paper IV**.

5.3.2 Numerical Illustration

For illustration, a MID of 0.05 is considered for utility values and regression coefficients of health state descriptors.

Approach 1 - To achieve desired precision of an estimated mean utility value for a particular health state

Suppose the study team is interested in calculating a sample size such that the predicted utility value of '12334' health state, which has $x_0'Dx_0$ value approximately equal to the mean of $x_0'Dx_0$ values (i.e., $\overline{D}_{x_0} = 0.210$) of all directly valued health states in the EQ-VT protocol, is not far from its true value by more than an MID for utility values. That is, the half-width of the 95% PI of the estimated utility value equal to δ = 0.05 (MID). The sample size can be calculated using equation (10), assuming $\hat{\sigma}^2 = 0.365$ and design effect $\overline{DE}_y$ = 0.75 (based on the Basic RE model), as $N = 10\,\hat{\sigma}^2\,\overline{D}_{x_0}\overline{DE}_y\,Z^2_{1-\alpha/2}/\delta^2$ = 10×0.365×0.210×0.75×(1.960)²/(0.05)² = 2.2084/0.0025 ≈ 883 participants.

The solid line in **Figure 8** shows the sample sizes calculated as illustrated above for Approach 1 for different MID (y-axis on the left). It shows a rapid decline in the curve with an increase in sample size from 100 to 400 participants. However, it decreases by a trivial magnitude with further increase of sample size after 400.

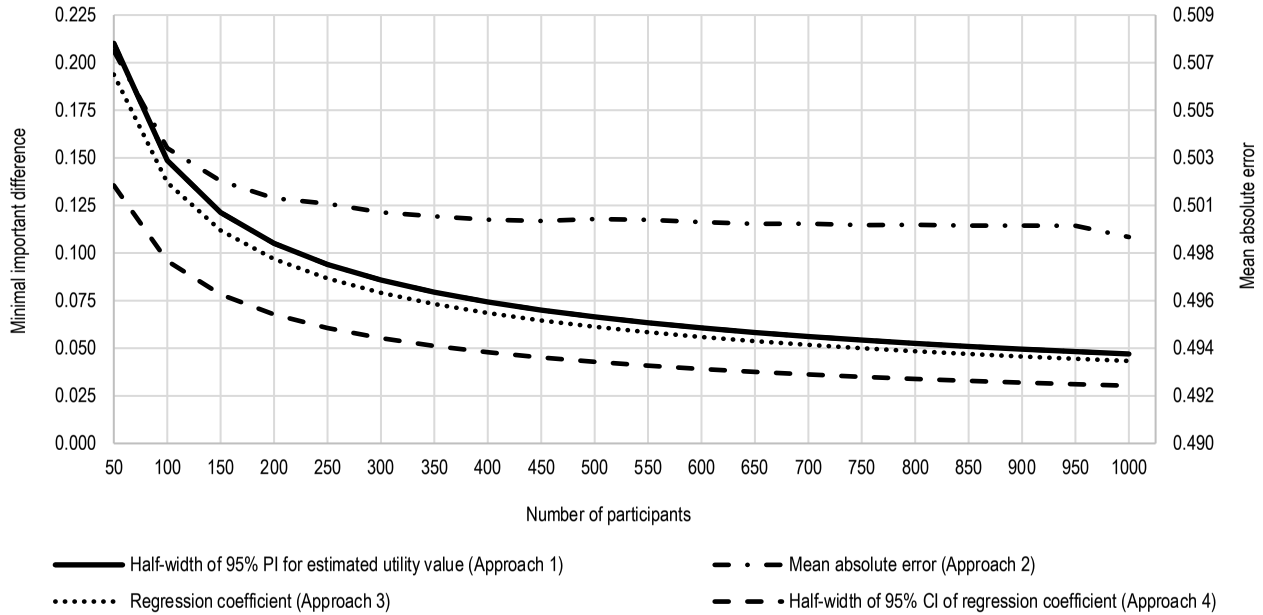


Figure 8. Comparison of sample sizes estimated using different approaches

PI Prediction interval, CI confidence interval. Refer to y-axis on the left for Approach 1, 3, and 4, and y-axis on the right for Approach 3. For Approach 2, mean absolute errors are based on 1000 replications for sample sizes from 100 to 900 participants (each of the participants valued 10 health states), whereas they are based on a single sample for the sample size of 1000 participants.

Approach 2 - Empirical approach to achieve desired mean absolute error in prediction of utility values relative to a reference study

For this approach, **Figure 8** shows the mean of MAE_{N_s} (y-axis on right) for 100 random samples (with replacement) of sample sizes 100 to 900 participants based on the Basic OLS model using the study C data (N = 1000). The MAE decreased rapidly between the sample sizes of 100 to 300. After that, reduction in the slope is trivial beyond the sample size of 300. That is, gain in prediction accuracy was trivial when the sample size was more than 300 participants.

Approach 3 - To assess significance of a regression coefficient of health state descriptors

The sample size required to test $H_0: \bar{\beta} = 0$ against $H_1: \bar{\beta} \neq 0$, assuming the true value of $\bar{\beta} = 0.05$ (MID) and $\hat{\sigma}^2$ = 0.365, at 5% two-sided level of significance (*a*) and 80% statistical power (1−*β*), can be calculated using equation (14), by inserting value of $(Z_{1-\alpha/2} + Z_{1-\beta})^2$ = (1.960 + 0.842)² = 7.851, $\bar{D} = 0.131$, and adjusting for design effect $\overline{DE}_{\beta} = 0.5$ (based on the Basic RE model), as $N = 10\,\hat{\sigma}^2\,\bar{D}\,\overline{DE}_{\beta}\,(Z_{1-\alpha/2} + Z_{1-\beta})^2/\bar{\beta}^2$ = 10×0.365×0.131×0.5×7.851/(0.05)² = 1.8770/0.0025 ≈ 751 participants. **Figure 8** shows sample sizes corresponding to values of $\bar{\beta}$ (y-axis on left) at 5% of α and 80% statistical power.

Approach 4 - To estimate a regression coefficient of health state descriptors with a desired precision

To estimate a regression coefficient with its 95% CI of half-width equal to $\bar{\delta} = 0.05$ (MID), using equation (17) with $\hat{\sigma}^2$ = 0.365, $\bar{D} = 0.131$ and $\overline{DE}_{\beta} = 0.5$ (as in previous illustration), the required sample size is $N = 10\,\hat{\sigma}^2\,\bar{D}\,\overline{DE}_{\beta}\,Z^2_{1-\alpha/2}/\bar{\delta}^2$ = 10×0.365×0.131×0.5×(1.960)²/(0.05)² = 0.9184/0.0025 ≈ 367 participants. **Figure 8** shows sample sizes corresponding to values of $\bar{\delta}$ (y-axis on left) for 95% CI of a regression coefficient.

The backward difference coding scheme used for health state descriptors has no impact on sample size estimation in Approach 1 and 2, and only a trivial impact on Approach 3 and 4 (Results not shown).

6 DISCUSSION

Several methodological issues in health state valuation were studied. Firstly, how the most severe state in the presence of a label, such as 'all-worst' is valued in a general population is explored using two of the most extensively utilized generic instruments, SF-6D and EQ-5D-3L, with the VAS method. This is the first study conducted to evaluate the impact of labeling a generic health state on its valuation. So far, all prior studies were conducted to study this aspect in disease-specific health states.(Green et al., 2017, Robinson and Bryan, 2013, Brazier et al., 2012, Rowen et al., 2012) Second, the impact of chronic disease experience on hypothetical health state valuation is studied in several patient populations using the EQ-5D-3L with the VAS method and the EQ-5D-5L with the composite TTO method. This issue was studied in two separate studies. The first study evaluated the impact of chronic disease experience on the valuation in patients with diabetes, rheumatism, hypertension, heart disease and lung disease. The study included a vast range of 42 EQ-5D-3L health states. This is the first study evaluating the impact of chronic diseases on the valuation of hypothetical health states with several chronic disease patients in the Asian population. The second study was a follow-up study to further evaluate the findings of the first study in the two high disease burden chronic disease populations, heart disease and cancer patients, with the latest revised version of EQ-5D-3L, EQ-5D-5L, using the standard protocol for the EQ-5D-5L value set based on the composite TTO method. In this study, the patients' diagnoses were also confirmed clinically. Finally, capitalizing on the standard valuation protocol for the EQ-5D-5L value set studies, statistical methods for estimating sample size for value set studies have been proposed. Previously used sample size methods in the EQ-5D-3L value set studies were either not based on the primary objective and analysis strategy for developing a value set or not capitalized on health states directly valued as per the valuation protocol.(Chevalier and de Pouvourville, 2013, Lamers et al., 2006, Dolan et al., 1994) My approaches will also help to evaluate the appropriateness of the currently recommended sample size of 1000 participants for the EQ-5D-5L value set studies. I hope my findings and recommendations on above methodological issues can contribute to improving health state valuation methods,

the scientific validity of value sets and guide selection of value sets for CUAs and medical decision making from a patient-centric approach.

6.1 Impact of Labeling on the Most Severe Health State Valuation

The most severe health state is one of the most important health states in value set studies. Usually, it is valued lowest among all the health states and determines the lower limit of the value set. It is often used to rescale utility values to get the value set with an appropriate range (e.g., 1 to -1).(Lamers, 2007) In this case, the value of the most severe health state has an impact on all the rescaled values. The majority of the value set studies include the most severe health state to be directly valued by each of the participants to increase the precision of its estimated value when considering its importance.(Szende et al., 2007) It is not uncommon to label the most severe health state such as 'all-worst' or 'pits' in value set studies.(Ratcliffe et al., 2016, Craig et al., 2013, Wee et al., 2006) However, an impact of such labeling on its valuation is not studied.

In my study, the most severe health state was labeled as 'all-worst.' I observed a deficit in the most severe state valuation which is unable to explain using its description (i.e., considering severity levels in each of the dimensions). This was, however, not observed in other least valued health states which are better by just severity level in one of the dimensions. The deficit in the valuation of the most severe health state was observed in valuation using both the commonly used health state utility instruments, EQ-5D-3L and SF-6D. The deficit was larger than the MID of the respective health state utility instruments as well as statistically significant, which indicates that the deficit has a practically significant impact.

One of the possible reasons for the deficit in the valuation of the most severe health state could that participants might have valued the most severe state based on the 'all-worst' label without fully based on its description according to the health state utility instrument. A few studies evaluating the impact of labeling on disease-specific health states in cancer, mental health and sclerosis also observed a lower valuation in the presence of labeling.(Green et al., 2017, Robinson and Bryan, 2013, Rabin et al., 1993) The current study showed a comparable labeling effect in a non-

disease specific health state and a general population context. Thus, we suggest avoiding the use of labeling health states in valuation studies.

Another possibility for the deficit in the valuation of the most severe health state could be due to end-aversion bias.(Torrance et al., 2001) That is, some participants could be systematically avoiding valuing any health state at the extreme end of the VAS. The VAS method involves deciding the least desirable health states from the most severe health state or dead state to be placed at the lower end of the VAS scale, and then rate the remaining health states between the perfect health the least valued health state. The health states near the lower end of the VAS scale might be valued at a distance from the most severe health state by the participants who consider the most severe health state worse than death, which could have created deficit in the valuation of the most severe health state. On the other hand, the participants who considered the most severe health state better than death, the end-aversion bias is not expected to affect the valuation of the most severe health state.

In my study, the deficit in the valuation of the most severe health state was more prominent in the valuation using the SF-6D compared to EQ-5D-3L. This can be explained by the differences in these two descriptive systems. A previous study has shown that participant considers severe EQ-5D-3L health states more severe than severe SF-6D health states.(Brazier et al., 2004) Therefore, the most severe health state of EQ-5D-3L is expected to be valued lower than the SF-6D counterpart, which leaves less room for labeling effect on its valuation.

6.2 Impact of Chronic Disease Experience on Health State Valuation

There is growing interest in how to incorporate patient preferences in the healthcare decision making, including in the HTAs for new treatments.(Underwood, 2016, Bridges and Jones, 2007) Some of the regulatory bodies recommend using the general population derived utility values for CUA aiming for reimbursement of the treatment using the societal resources.(Sanders et al., 2016, NICE, 2013) However, the targeted patient population derived utility values are relevant for medical decision making from a patient-centric approach.(Stamuli, 2011) Nevertheless, whose utility values should be used depends on whether the utility values derived by the general population and the patient population differ significantly. There are only a few

studies available comparing the utility of hypothetical health states between patient populations and the general population.(Ogorevc et al., 2017, Wang et al., 2014, Pickard et al., 2013, Krabbe et al., 2011) Furthermore, these studies have shown mixed findings. I compared the utility values of hypothetical health states described by the most widely used generic health state utility instrument, EQ-5D, between several patient populations and the general population (potentially without chronic diseases) in two separate studies.

In the first study (study A), the mean value of EQ-5D-3L health states based on the VAS method was compared between participants with self-reported chronic diseases (diabetes, rheumatism, hypertension, heart disease and lung disease) and participants without any chronic diseases. The strength of the study was that it used a vast range of 42 health states for valuation and included several chronic disease patient populations.

The study showed that heart disease patients value hypothetical health states, mainly severe health states, higher than participants with no chronic disease. The mean difference in the values between heart disease patients and participants with no chronic disease was statistically significant and also larger than the MID for EQ-5D-3L values. This reflects the practically significant impact of the heart disease experience on the health state valuation. No such difference was observed in the valuation of other chronic diseases - diabetes, rheumatism, hypertension, and lung disease - and participants with no chronic disease.

In the second study (study B), similar to study A, I compared the mean value of hypothetical health states described by the revised version of EQ-5D-3L, EQ-5D-5L, valued using the currently recommended valuation method - the composite TTO method, by patients with heart disease and cancer with participants from the general population. The strength of the study was that patient diagnosis was medically confirmed, a fixed set of health states were valued by patients and the general population participants, the same interviewer conducted all the interviews (to minimize interviewer effect).

The study showed that heart disease patients value hypothetical health states, mainly mild health states, higher than participants from the general population. The mean difference in the values between heart disease patients and participants from the general population was statistically significant and also larger than the MID for EQ-5D-5L values. This reflects the practically significant impact of the heart disease

experience on the health state valuation. No such difference was observed in the valuation of patients with cancer and participants from the general population.

A post-hoc analysis of study A, comparing values of mild health states between patients with heart disease and participants with no chronic disease, also showed that heart disease patients value mild health states statistically significantly higher than the no chronic disease participants, and the difference was larger than the MID for the EQ-5D-3L (see **Appendix 7** for more details). This concurs with the findings of study B.

A possible reason for heart disease patients giving higher valuation scores could be that a more significant proportion of heart disease patients might have experienced health states of varying severity, and this might have changed their perception regarding the severity of health problems. That is, heart disease patients are more likely to have adapted to health issues demarcated by EQ-5D states than healthy ones. Consequently, patients do not distinguish those health states as unbearable or unwanted as the general population. This might not be the case with other chronic disease patients, or it could at least be various levels of adaptation.

There is another conceivable reason for why heart disease and other chronic disease patients, such as cancer patients, were dissimilar in the valuation of EQ-5D health states in contrast to the general population. It is possible that, as opposed to the general population and patients with heart disease of a similar age, those with cancer are more prone to perceiving their actual life expectancy to be less than 10 years and are consequently more generous when trading the 'extra' years. This tendency may lead to lower utility values for all severities of health states. It could offset the result of adaptation and as a consequence, make patients with cancer similar to the general population in the valuation of health states using the TTO method. The finding that older age is associated with lower utility value in study B may be evidence for the effect of self-professed life expectancy on TTO-based valuation using a fixed time frame.

The absence of meaningful differences in the valuation between some chronic disease patients and the general population might also be because several hypothetical health states included in my studies might not be actually experienced by patients in real life. Hence, they value these health states similar to participants with no chronic disease.

The findings from study B are consistent with the study reported by Pickard *et al.*(Pickard et al., 2013) based on the TTO values for EQ-5D-3L health states. Pickard *et al.* reported no meaningful difference in utility values between patients with chronic diseases (arthritis, diabetes, depression, hay fever, cancer) and participants with no chronic disease, except for heart failure patients. The study showed that patients with heart failure only, and patients with heart failure and at least one other chronic disease, gave values higher by 0.25 points and 0.07 points, respectively, which is larger than the MID for EQ-5D-3L, compared to participants with no chronic disease.

In another study conducted by Krabbe *et al.*(Krabbe et al., 2011) comparing VAS and TTO values for EQ-5D-3L health states between patients with rheumatoid arthritis and cancer with participants from the general population, showed no meaningful difference in VAS values between patients with rheumatoid arthritis and participants from the general population. However, he showed that cancer patients give higher TTO values compared to participants from the general population. It should be noted that this study did not adjust for any sociodemographic characteristics, including age. Patients in this study were much older (mean age 63 years and 65 years for cancer and rheumatoid arthritis patients, respectively) compared to participants from the general population (mean age 44 years). As age can also affect the valuation (shown in my study B), the results of this study are not fully comparable with my study results.

6.3 Determination of Sample Size for Value Set Studies

The standard valuation protocol, EQ-VT, based on the composite TTO method for the EQ-5D-5L value sets recommends a sample size of 1000 participants.(Oppe et al., 2014) The recommended sample size lacks sound statistical justification as well as empirical evidence for its appropriateness. The determination of a suitable sample size is very important: a sample size that is too small will not allow achieving the research purpose, whereas a sample size that is too large will compromise feasibility of the study. I propose four approaches for determining the sample size for the EQ-5D-5L value set studies according to the standardized protocol.

Four approaches for determining sample size were proposed and illustrated for the valuation studies of health states using EQ-VT protocol—the first two are based

on achieving desired precision in predicting utility values, and remaining two on regression coefficients for a model predicting utility values. Useful sample size parameters are estimated from a Singaporean study. Amid the first two approaches which are directly associated with the primary aim of value set studies to predict utility values with desired precision, Approach 1 (based on prediction intervals of predicted utility values) has certain advantages over Approach 2 (based on mean absolute errors in predicted utility values). In the Approach 1, the study team can choose the desired level of precision (prediction error) with the desired level of confidence (probability), whereas in Approach 2 is based on achieving the mean level of precision (prediction error). Moreover, it also requires data from a similar value set study with a larger sample size and the sample size of a new study be selected proportionate to the larger study. The remaining two approaches are also associated with the analysis plan for developing a value set with acceptable precision.

Previous sample size estimation methods used in the EQ-5D-3L value set studies were based on comparing mean utility values of two directly valued health states or estimating utility values of directly valued health states with the desired precision (confidence interval)(Chevalier and de Pouvourville, 2013, Lamers et al., 2006, Dolan et al., 1994). These methods have either not incorporated regression model used to predict utility values or capitalized the method based on the directly valued health states. In other words, these sample size approaches were not directly based on the study's primary objective, analysis plan or protocol. Hence, the precision of predicted utility values using these sample size methods cannot be assured. An approach based on the study protocol, primary aim, outcome, and analysis plan should be chosen by the study team.

Two approaches commonly used in clustered data for statistical inference - cluster-robust estimation and random-effects model – were studied to evaluate how they influence the sample size estimation. Both methods exhibited deflation in the variances for regression coefficients of model predicting utility values when adjusted for within-person correlation. This was expected as each block involves health states from varying severity from mild to severe; as such, the within-person correlation among allocated health states is negative.(Kahan and Morris, 2013, Parzen et al., 1998) Apropos the choice between the clustered-robust estimator and random-effects model methods for estimating the design effect, is not straightforward and rests on the preference of the study team for the model selection and also on data. The approach of the cluster-robust estimator needs fewer assumptions in contrast

to that of the random-effects model approach.(Cameron and Miller, 2015) Nonetheless, it is less efficient. Better efficiency is seen with the random-effects model which gives smaller standard errors and design effect estimate contrary to the other approach. However, there is no assurance that the model assumptions in the random-effects model hold for given data. Nonetheless, as is conveyed in the Results section, the Pearson's correlation between utility values predicted using the ordinary least-square model (with cluster-robust standard errors) and the random-effects models in the Singapore study, was close to 1, signifying both models provided almost identical results.

6.4 Limitations

There are several limitations in my studies. All my studies were conducted in Singapore. As the cultural differences affect the health state valuation, the generalizability of the findings is limited. Nonetheless, as the Singapore is a multi-ethnic country with three major ethnicities – Chinese, Malay, and Indian. The study findings have potential to generalize in the Asian population. Further research will be required in western countries to generalize the findings in the western population.

The main limitation of the study evaluating the labeling effect on the most severe health state was that the labeling effect was evaluated in the absence of a control arm, without a label for the most severe health state. Therefore, the findings of the study required to be confirmed in a randomized controlled study with and without labeling arms. This study used the VAS method for valuation. As the valuation method could impact the health state valuation, further studies will be required to evaluate the labeling effect using other valuation methods.

Both the studies comparing valuation using the EQ-5D-3L (study A) and EQ-5D-5L (study B) have a few common limitations. First, the study samples did not incorporate inpatients or patients at a severe or unstable stage of chronic diseases. As such, inpatients are likely to be in the worse functional state and might be associated with a lower valuation. For example, heart disease outpatients in the EQ-5D-5L study (study B), whose functional state was associated with a lower valuation (data not shown). Therefore, the difference in valuation between chronic disease patients and the general population could be under-estimated. Nonetheless, it is challenging to conduct the highly demanding cognitive valuation interviews with

extremely severe patients. Second, the studies involved only two life-threatening chronic diseases (cardiovascular diseases and cancers). Further studies in other life-threatening chronic conditions will be required to generalize the findings to a wider group of chronic conditions. Third, separate statistical tests were performed for comparing valuation of health states with varying severity between each of the patient populations and the control population without multiplicity adjustment. The studies were not powered for these analyses. Therefore, the statistical significance observed in these studies might be due to inflated Type-I error probability and findings require further confirmation. Lastly, because of the sample size constraint, it was impossible to evaluate the effect of chronic disease experience on the valuation of individual dimensions of EQ-5D health states.

In the study comparing the EQ-5D-3L valuation (study A), the chronic diseases were self-reported by participants. Therefore, I cannot rule out the possibility of misclassification as they were not clinically confirmed. Nevertheless, there could be less possibility of participants with no disease to report having a disease than patients with a disease reporting having no disease. A study conducted in Finnish population(Oksanen et al., 2010) showed that the sensitivity of self-reporting chronic conditions such as diabetes, hypertension, heart disease, asthma and rheumatoid arthritis could range from 78 to 96%, whereas the specificity could range from 96 to 99%. That is, the difference in valuation between patients and participants with no chronic disease might be under-estimated, but fewer chances that it is over-estimated.

In the study comparing the EQ-5D-5L valuation (study B), the general population sample was enlisted from shopping malls. Although malls are a commonly visited place by the Singaporeans(URA, 2009), shopping is typically an enjoyable leisure activity which might influence the valuation of health states. Therefore, the extent of the difference observed in valuation by patients and general population could also be affect by this sample selection.

It should be noted that while developing methodologies for sample size estimation for EQ-5D-5L value set studies, my objective was not to recommend any analysis strategy or model, nor offer a value set. Therefore, no attempt to optimize the model fitting has been made, except for incorporating 20 indicator variables and an intercept intended for health state descriptors. No interaction terms have been included in these basic models, which could be valuable for the model assumptions

and performance validity. It could also be the likely reason why when using Approach 2, the mean absolute error has a high value, even when the sample size of participants is 1000. By what method the analysis models can be improved and the prediction error to be reduced, is a complex problem beyond the scope of the current research, and in theory, is associated with valuation methodology in addition to modeling technique. This may necessitate additional research. Moreover, when illustrating the sample size determination approaches, a minimally important difference of 0.05 was used for utility values simply with the intent to illustrate the numerical applications. In certain realistic circumstances, for example, the illustration in Approaches 1 and 3, 1000 participants as a sample size was more or less right. Nevertheless, they were based on certain assumptions and a particular level of precision, which could differ according to the study team's requirements. The sample size should be determined by the study team as per its requirements.

7 SCIENTIFIC CONCLUSIONS

1. The most severe state with label 'all-worst' is valued significantly lower than expected, according to its descriptors. Labels in addition to health state description should be avoided to minimize the impact of unintended factors on health state valuation. (**Publication Paper I**)

2. Chronic disease patients might vary from those of the general population when it came to the strength of their preferences for hypothetical health states. The difference in preferences may depend on the type of the disease, the patient's age, and how severe the health state being valued is. Consequently, using utility values originated from the general population might undervalue or overvalue the comparative effectiveness of healthcare interventions for specific kinds of diseases, for example, heart diseases. (**Publication Papers II and III**)

3. The proposed four sample size estimation approaches and the relevant parameter estimates can help to decide an appropriate sample size for a value set study. (**Publication Paper IV**)

8 IMPLICATIONS AND FUTURE RESEARCH NEEDS

The present research findings provide several recommendations for improvement in health state valuation techniques and selection of value sets in CUA and comparative effectiveness from a patient-centric approach.

The findings based on the first research aim suggest avoiding using any labels (e.g., 'all-worst' for the most severe health state) in addition to health state description. In the presence of labeling, the valuation of health states might be affected by prior belief or emotions associated with the label. Specifically, if a label is attached to the anchoring health states, such as the most severe health state which often decides lower bound of utility value sets, may have major implications. As the current study was a single-arm observational study with all participants exposed to a health state with a label, a randomized controlled study with participants exposed to health states with and without a label will be needed to confirm the findings.

The findings from the second research aims indicated that some specific chronic disease experience might affect how individuals with such experience perceive different health conditions. For example, my studies showed patients with heart disease experience may not consider a health condition with mild impairment as worse compared to individuals with no such experience may consider. Similarly, severe health conditions were also perceived differently by heart disease patients. Such systemic differences in perception might lead to differences in health state valuation derived from the patient and the general populations, which could ultimately lead to under-estimation or over-estimation of the effectiveness of a healthcare intervention if the general population derived value set is used. This finding may have implications in clinical decision making as to which healthcare intervention is preferable for patients with the disease. If a patient-centered approach is to be adopted to guide patient care to ensure that targeted patient population preferences and values are incorporated, the patient-derived utility value set should be considered. As such, patient-derived utility value set will be needed for

the evaluative effectiveness of available healthcare interventions. As our studies were limited to a few chronic diseases and limited sample sizes, further studies with larger sample sizes and patients with different chronic diseases and severity should be conducted to understand the implication of disease experience on health state valuation. If the patient valuation is found to be systematically different from the general population, a utility set targeted to the specific patient population should be developed for patient-centered clinical decision-making and effectiveness analysis.

Lastly, the third research aim suggested multiple approaches for estimating required sample size for EQ-5D-5L value set studies. These approaches were capitalized on the latest internationally adopted protocol for EQ-5D-5L health state valuation. The suggested approaches can help to determine sample size for a country-specific value set for the general population as well as a value set specific to a particular patient population. The sample size approaches offer to determine sample sizes using given levels of precision with different objectives related to the development of value set studies. Further research will be needed to generalize the sample size approaches for other multi-attribute descriptive systems with different valuation methodologies.

9 ACKNOWLEDGEMENTS

I would like to thank the Tampere Center for Child Health Research at the University of Tampere and Tampere University Hospital, Duke-NUS Medical School, National University of Singapore, National Heart Centre Singapore, Singapore General Hospital, and National Cancer Centre Singapore for their support in my study. Special thanks to my current employer Singapore Clinical Research Institute for allowing me to pursue my study and supporting my research in various aspects.

Some of the coursework was done at the University of Tampere and the National University of Singapore. I thank the facilitators for giving me the opportunity to learn important topics related to my research.

The methodological research was carried out in the Tampere Center for Child Health Research at the University of Tampere and Tampere University Hospital, while empirical studies were conducted in the National Heart Centre Singapore, National Cancer Centre Singapore and at various communities (residential areas) in Singapore. I would like to express my heartfelt gratitude to my all collaborators for their great support at every phase of the research. However, the following people deserve special mention:

Professor Yin Bun Cheung, for his overall supervision and guidance throughout my studies and being my mentor. Yin Bun introduced me to the health-related quality of life research, study designs, scientific thinking, and writing. These helped me a lot to develop myself as a researcher from a statistician. His advice was not limited to this research but to also successes in my professional development. I am very grateful for his time and guidance. He is indeed a role model for me.

Professor Julian Thumboo, for co-supervision my studies. His guidance and encouragement were invaluable in completing this work. His expertise and guidance in conducting research studies in patient population were very helpful. I am also very

thankful to him for allowing to use his study data for my research, without which it could have not possible to do this research.

Professor Nan Luo, National University of Singapore and Professor Ru San Tan, National Heart Centre Singapore for their support regarding providing manpower for conducting studies. Their vast knowledge and experience benefited me to a great extent in planning, conducting my research studies and interpreting their results.

Professor Per Ashorn, for reviewing my research progress and encouragement throughout my study. I am also thankful for his guidance in facilitating my study without being in Tampere full-time.

Faculty and staff from the Tampere Center for Child Health Research, Tampere University. Special thanks to Kirsi-Maarit Lehto for administrative support.

I thank the assistance of Grace Yang and Qu Limin (Debra) in collecting cancer patient data. Also, thanks to Yeoh Yen Shing for the assistance in data management.

All the study participants and interviewers. Without them, this study would not have been possible.

All the co-authors who collaborated in the original manuscripts that contributed to the thesis. Your advice, encouragement and timely comments were always greatly appreciated.

Thanks to my immediate manager, Professor Edwin Chan and other colleagues in Singapore for your friendship and support for the Ph.D.

My dear mother and sisters for their cheering for me. Lastly, my beloved wife for being the strength of my life and sharing the ups and downs of my life!

Singapore, March 2018

Mihir Gandhi

10 REFERENCES

AL SAYAH, F., MLADENOVIC, A., GAEBEL, K., XIE, F. & JOHNSON, J. A. 2016. How dead is dead? Qualitative findings from participants of combined traditional and lead-time time trade-off valuations. *Quality of Life Research,* 25**,** 35-43.

BANSBACK, N., HOLE, A. R., MULHERN, B. & TSUCHIYA, A. 2014. Testing a discrete choice experiment including duration to value health states for large descriptive systems: Addressing design and sampling issues. *Social Science & Medicine,* 114**,** 38-48.

BRAUER, C. A., ROSEN, A. B., GREENBERG, D. & NEUMANN, P. J. 2006. Trends in the measurement of health utilities in published cost-utility analyses. *Value in Health,* 9**,** 213-218.

BRAZIER, J., ARA, R., ROWEN, D. & CHEVROU-SEVERAC, H. 2017a. A review of generic preference-based measures for use in cost-effectiveness models. *PharmacoEconomics,* 35**,** 21-31.

BRAZIER, J., RATCLIFFE, J., SALOMAN, J. & TSUCHIYA, A. 2017b. *Measuring and valuing health benefits for economic evaluation,* New York, OXFORD University Press.

BRAZIER, J., ROBERTS, J. & DEVERILL, M. 2002. The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics,* 21**,** 271-292.

BRAZIER, J., ROBERTS, J., TSUCHIYA, A. & BUSSCHBACH, J. 2004. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics,* 13**,** 873-884.

BRAZIER, J., ROWEN, D., MAVRANEZOULI, I., TSUCHIYA, A., YOUNG, T., YANG, Y., BARKHAM, M. & IBBOTSON, R. 2012. 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,* 16**,** 1-114.

BRIDGES, J. F. & JONES, C. 2007. Patient-based health technology assessment: a vision of the future. *International Journal of Technology Assessment in Health Care,* 23**,** 30-35.

CAMERON, A. C. & MILLER, D. L. 2015. A practitioner's guide to cluster-robust inference. *Journal of Human Resources,* 50**,** 317-372.

CHEN, X., ENDER, P., MITCHELL, M. & WELLS, C. 2003. *Regression with STATA* [Online]. Available: https://stats.idre.ucla.edu/stata/webbooks/reg/ [Accessed 28th February 2018].

CHEVALIER, J. & DE POUVOURVILLE, G. 2013. Valuing EQ-5D using time trade-off in France. *The European Journal of Health Economics,* 14**,** 57-66.

CORETTI, S., RUGGERI, M. & MCNAMEE, P. 2014. The minimum clinically important difference for EQ-5D index: a critical review. *Expert Review of Pharmacoeconomics & Outcomes Research,* 14**,** 221-233.
CRAIG, B. M., PICKARD, A. S., STOLK, E. & BRAZIER, J. E. 2013. US valuation of the SF-6D. *Medical Decision Making,* 33**,** 793-803.
DEVLIN, N. J. & BROOKS, R. 2017. EQ-5D and the EuroQol Group: past, present and future. *Applied Health Economics and Health Policy,* 15**,** 127-137.
DEVLIN, N. J., SHAH, K. K., FENG, Y., MULHERN, B. & HOUT, B. 2017. Valuing health-related quality of life: An EQ-5D-5L value set for England. *Health Economics,* 27**,** 7-22.
DOLAN, P. 1997. Modeling valuations for EuroQol health states. *Medical Care,* 35**,** 1095-1108.
DOLAN, P., GUDEX, C., KIND, P. & WILLIAMS, A. 1994. *The measurement and valuation of health: first report on the main survey* [Online]. Available: https://www.york.ac.uk/media/che/documents/reports/MVH%20First%20Report.pdf [Accessed 28th February 2018].
DONNER, A. & KLAR, N. 2000. *Design and analysis of cluster randomization trials in health research,* New York, Oxford University Press.
DOS. 2010. *Census of population 2010 statistical release 1. Demographic characteristics, education, language and religion* [Online]. Singapore Department of Statistics. Available: https://www.singstat.gov.sg/docs/default-source/default-document-library/publications/publications_and_papers/cop2010/census_2010_release1/cop2010sr1.pdf [Accessed 28th February 2018].
FENG, Y., DEVLIN, N. J., SHAH, K. K., MULHERN, B. & HOUT, B. 2017. New methods for modelling EQ-5D-5L value sets: An application to English data. *Health Economics,* 27**,** 23-38.
GREEN, C., BRAZIER, J. & DEVERILL, M. 2000. Valuing health-related quality of life. *Pharmacoeconomics,* 17**,** 151-165.
GREEN, C., GOODWIN, E. & HAWTON, A. 2017. "Naming and framing": the impact of labeling on health state values for multiple sclerosis. *Medical Decision Making,* 37**,** 703-714.
GUDEX, C., DOLAN, P., KIND, P. & WILLIAMS, A. 1996. Health state valuations from the general public using the visual analogue scale. *Quality of Life Research,* 5**,** 521-531.
IKEDA, S., SHIROIWA, T., IGARASHI, A., NOTO, S., FUKUDA, T., SAITO, S. & SHIMOZUMA, K. 2015. Developing a Japanese version of the EQ-5D-5L value set. *J Natl Inst Public Health,* 64**,** 47-55.
ISPOR. 2017. *Pharmacoeconomic guidelines around the world* [Online]. ISPOR. Available: https://tools.ispor.org/peguidelines/ [Accessed 8th September 2018].
JOHNSON, J. A., LUO, N., SHAW, J. W., KIND, P. & COONS, S. J. 2005. Valuations of EQ-5D health states: are the United States and United Kingdom different? *Medical Care,* 43**,** 221-228.
KAHAN, B. C. & MORRIS, T. P. 2013. Assessing potential sources of clustering in individually randomised trials. *BMC Medical Research Methodology,* 13**,** 58.

KARIMI, M., BRAZIER, J. & PAISLEY, S. 2017a. Are preferences over health states informed? *Health and Quality of Life Outcomes,* 15**,** 105.
KARIMI, M., BRAZIER, J. & PAISLEY, S. 2017b. How do individuals value health states? A qualitative investigation. *Social Science & Medicine,* 172**,** 80-88.
KIM, S.-H., AHN, J., OCK, M., SHIN, S., PARK, J., LUO, N. & JO, M.-W. 2016. The EQ-5D-5L valuation study in Korea. *Quality of Life Research,* 25**,** 1845-52.
KRABBE, P. F., TROMP, N., RUERS, T. J. & VAN RIEL, P. L. 2011. Are patients' judgments of health status really different from the general population? *Health and Quality of Life Outcomes,* 9**,** 31.
KWAN, Y. H., FONG, W. W. S., LUI, N. L., YONG, S. T., CHEUNG, Y. B., MALHOTRA, R., ØSTBYE, T. & THUMBOO, J. 2016. Validity and reliability of the short form 36 health surveys (SF-36) among patients with spondyloarthritis in Singapore. *Rheumatology International,* 36**,** 1759-1765.
LAMERS, L. M. 2007. The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value sets. *Medical Care,* 45**,** 238-244.
LAMERS, L. M., MCDONNELL, J., STALMEIER, P. F., KRABBE, P. F. & BUSSCHBACH, J. J. 2006. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Economics,* 15**,** 1121-1132.
LUDWIG, K., VON DER SCHULENBURG, J.-M. G. & GREINER, W. 2018. German Value Set for the EQ-5D-5L. *PharmacoEconomics,* (Epub).
LUO, N., JOHNSON, J. A. & COONS, S. J. 2010. Using instrument-defined health state transitions to estimate minimally important differences for four preference-based health-related quality of life instruments. *Medical Care,* 48**,** 365-371.
LUO, N., LIU, G., LI, M., GUAN, H., JIN, X. & RAND-HENDRIKSEN, K. 2017. Estimating an EQ-5D-5L value set for China. *Value in Health,* 20**,** 662-669.
LUO, N., WANG, P., THUMBOO, J., LIM, Y.-W. & VRIJHOEF, H. J. 2014. Valuation of EQ-5D-3L health states in Singapore: modeling of time trade-off values for 80 empirically observed health states. *Pharmacoeconomics,* 32**,** 495-507.
LUO, N., WANG, Q., FEENY, D., CHEN, G., LI, S.-C. & THUMBOO, J. 2007. Measuring health preferences for Health Utilities Index Mark 3 health states: a study of feasibility and preference differences among ethnic groups in Singapore. *Medical Decision Making,* 27**,** 61-70.
MALHOTRA, C., FAROOQUI, M. A., KANESVARAN, R., BILGER, M. & FINKELSTEIN, E. 2015. Comparison of preferences for end-of-life care among patients with advanced cancer and their caregivers: a discrete choice experiment. *Palliative Medicine,* 29**,** 842-850.
MCCLURE, N. S., AL SAYAH, F., XIE, F., LUO, N. & JOHNSON, J. A. 2017. Instrument-defined estimates of the minimally important difference for EQ-5D-5L index scores. *Value in Health,* 20**,** 644-650.
MCTAGGART-COWAN, H. 2011. Elicitation of informed general population health state utility values: a review of the literature. *Value in Health,* 14**,** 1153-1157.

MOH. 2010. *Singapore burden of disease study 2010* [Online]. Ministry of Health Singapore. Available: https://www.moh.gov.sg/content/dam/moh_web/Publications/Reports/2014/Singapore%20Burden%20of%20Disease%20Study%202010%20Report_v3.pdf [Accessed 28th February 2018].

MONTGOMERY, D. C., PECK, E. A. & VINING, G. G. 2012. *Introduction to linear regression analysis,* New Jersey, John Wiley & Sons.

NICE. 2013. *Guide to the methods of technology appraisal 2013* [Online]. National Institute for Health and Clinical Excellence. Available: https://www.nice.org.uk/process/pmg9/chapter/1-foreword [Accessed 28th February 2018].

NOLAN, C. M., LONGWORTH, L., LORD, J., CANAVAN, J. L., JONES, S. E., KON, S. S. & MAN, W. D. 2016. The EQ-5D-5L health status questionnaire in COPD: validity, responsiveness and minimum important difference. *Thorax,* 71**,** 493-500.

OGOREVC, M., MUROVEC, N., FERNANDEZ, N. B. & RUPEL, V. P. 2017. Questioning the differences between general public vs. patient based preferences towards EQ-5D-5L defined hypothetical health states. *Health Policy,* (Epub).

OKSANEN, T., KIVIMÄKI, M., PENTTI, J., VIRTANEN, M., KLAUKKA, T. & VAHTERA, J. 2010. Self-report as an indicator of incident disease. *Annals of Epidemiology,* 20**,** 547-554.

OPPE, M., DEVLIN, N. J., VAN HOUT, B., KRABBE, P. F. & DE CHARRO, F. 2014. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value in Health,* 17**,** 445-453.

OPPE, M., RAND-HENDRIKSEN, K., SHAH, K., RAMOS-GOÑI, J. M. & LUO, N. 2016. EuroQol protocols for time trade-off valuation of health outcomes. *PharmacoEconomics,* 34**,** 993-1004.

PARZEN, M., LIPSITZ, S. R. & DEAR, K. B. 1998. Does clustering affect the usual test statistics of no treatment effect in a randomized clinical trial? *Biometrical Journal,* 40**,** 385-402.

PEETERS, Y. & STIGGELBOUT, A. M. 2010. Health state valuations of patients and the general public analytically compared: a meta-analytical comparison of patient and population health state utilities. *Value in Health,* 13**,** 306-309.

PICKARD, A. S., TAWK, R. & SHAW, J. W. 2013. The effect of chronic conditions on stated preferences for health. *The European Journal of Health Economics,* 14**,** 697-702.

PURBA, F. D., HUNFELD, J. A., ISKANDARSYAH, A., FITRIANA, T. S., SADARJOEN, S. S., RAMOS-GOÑI, J. M., PASSCHIER, J. & BUSSCHBACH, J. J. 2017. The Indonesian EQ-5D-5L value set. *PharmacoEconomics,* 35**,** 1153-1165.

RABIN, R., ROSSER, R. M. & BUTLER, C. 1993. Impact of diagnosis on utilities assigned to states of illness. *Journal of the Royal Society of Medicine,* 86**,** 444-448.

RAMOS-GOÑI, J. M., PINTO-PRADES, J. L., OPPE, M., CABASÉS, J. M., SERRANO-AGUILAR, P. & RIVERO-ARIAS, O. 2017. Valuation and

modeling of EQ-5D-5L health states using a hybrid approach. *Medical Care,* 55**,** e51-e58.
RAND-HENDRIKSEN, K., RAMOS-GOÑI, J. M., AUGESTAD, L. A. & LUO, N. 2017. Less is more: cross-validation testing of simplified nonlinear regression model specifications for EQ-5D-5L health state values. *Value in Health,* 20**,** 945-952.
RATCLIFFE, J., HUYNH, E., CHEN, G., STEVENS, K., SWAIT, J., BRAZIER, J., SAWYER, M., ROBERTS, R. & FLYNN, T. 2016. Valuing the Child Health Utility 9D: using profile case best worst scaling methods to develop a new adolescent specific scoring algorithm. *Social Science & Medicine,* 157**,** 48-59.
ROBINSON, S. & BRYAN, S. 2013. Does the process of deliberation change individuals' health state valuations? An exploratory study using the person trade-off technique. *Value in Health,* 16**,** 806-813.
ROWEN, D., BRAZIER, J., TSUCHIYA, A., YOUNG, T. & IBBOTSON, R. 2012. It's all in the name, or is it? The impact of labeling on health state values. *Medical Decision Making,* 32**,** 31-40.
RUSSELL, L. B., GOLD, M. R., SIEGEL, J. E., DANIELS, N. & WEINSTEIN, M. C. 1996. The role of cost-effectiveness analysis in health and medicine. *JAMA,* 276**,** 1172-1177.
SANDERS, G. D., NEUMANN, P. J., BASU, A., BROCK, D. W., FEENY, D., KRAHN, M., KUNTZ, K. M., MELTZER, D. O., OWENS, D. K. & PROSSER, L. A. 2016. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. *JAMA,* 316**,** 1093-1103.
SONNENBERG, F. A. & BECK, J. R. 1993. Markov models in medical decision making: a practical guide. *Medical Decision Making,* 13**,** 322-338.
STAMULI, E. 2011. Health outcomes in economic evaluation: who should value health? *British Medical Bulletin,* 97**,** 197-210.
SUAREZ-ALMAZOR, M. E. & CONNER-SPADY, B. 2001. Rating of arthritis health states by patients, physicians, and the general public. Implications for cost-utility analyses. *The Journal of Rheumatology,* 28**,** 648-656.
SZENDE, A., OPPE, M. & DEVLIN, N. 2007. *EQ-5D value sets: inventory, comparative review and user guide,* Dordrecht, Springer Science & Business Media.
THUMBOO, J., FONG, K.-Y., MACHIN, D., CHAN, S.-P., LEONG, K.-H., FENG, P.-H., THIO, S.-T. & BOEY, M.-L. 2001. A community-based study of scaling assumptions and construct validity of the English (UK) and Chinese (HK) SF-36 in Singapore. *Quality of Life Research,* 10**,** 175-188.
THUMBOO, J., WU, Y., TAI, E.-S., GANDEK, B., LEE, J., MA, S., HENG, D. & WEE, H.-L. 2013. Reliability and validity of the English (Singapore) and Chinese (Singapore) versions of the Short-Form 36 version 2 in a multi-ethnic urban Asian population in Singapore. *Quality of Life Research,* 22**,** 2501-2508.
TORRANCE, G. W., FEENY, D. & FURLONG, W. 2001. Visual analog scales: do they have a role in the measurement of preferences for health states? *Medical Decision Making,* 21**,** 329-34.

UNDERWOOD, G. 2016. *NICE to research patient preferences in HTA* [Online]. PharmaTimes. Available: http://www.pharmatimes.com/news/nice_to_research_patient_preferences_in_hta_1033029 [Accessed 28th February 2018].
URA. 2009. *The URA lifestyle survey 2009* [Online]. Urben Redevelopment Authority Singapore. Available: https://www.ura.gov.sg/-/media/User%20Defined/URA%20Online/Concept%20Plan/Public-Consulations/URA-Lifestyle-Survey-2009.pdf?la=en [Accessed 28th February 2018].
VERBEKE, G. 2000. *Linear mixed models for longitudinal data.* New York: Springer.
VERSTEEGH, M. M., VERMEULEN, K. M., EVERS, S. M., DE WIT, G. A., PRENGER, R. & STOLK, E. A. 2016. Dutch tariff for the five-level version of EQ-5D. *Value in Health,* 19**,** 343-352.
WALTERS, S. J. & BRAZIER, J. E. 2003. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health and Quality of Life Outcomes,* 1**,** 4.
WANG, P., TAI, E., THUMBOO, J., VRIJHOEF, H. J. & LUO, N. 2014. Does diabetes have an impact on health-state utility? A study of asians in singapore. *The Patient-Patient-Centered Outcomes Research,* 7**,** 329-337.
WARE JR, J. E. & SHERBOURNE, C. D. 1992. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical Care,* 30**,** 473-483.
WEE, H.-L., LI, S.-C., XIE, F., ZHANG, X.-H., LUO, N., CHEUNG, Y.-B., MACHIN, D., FONG, K.-Y. & THUMBOO, J. 2006. Are Asians comfortable with discussing death in health valuation studies? A study in multi-ethnic Singapore. *Health and Quality of Life Outcomes,* 4**,** 93.
WEE, H.-L., MACHIN, D., LOKE, W.-C., LI, S.-C., CHEUNG, Y.-B., LUO, N., FEENY, D., FONG, K.-Y. & THUMBOO, J. 2007. Assessing differences in utility scores: a comparison of four widely used preference-based instruments. *Value in Health,* 10**,** 256-265.
WHO. 2014. *World Health Statistics* [Online]. WHO Press. Available: http://apps.who.int/iris/bitstream/10665/112738/1/9789240692671_eng.pdf [Accessed 28th February 2018].
WIKIPEDIA. 2017. *Singapore* [Online]. Wikipedia. Available: https://en.wikipedia.org/wiki/Singapore [Accessed 28th February 2018].
WILLIAMS, R. L. 2000. A note on robust variance estimation for cluster-correlated data. *Biometrics,* 56**,** 645-646.
WISLØFF, T., HAGEN, G., HAMIDI, V., MOVIK, E., KLEMP, M. & OLSEN, J. A. 2014. Estimating QALY gains in applied studies: a review of cost-utility analyses published in 2010. *PharmacoEconomics,* 32**,** 367-375.
XIE, F., GAEBEL, K., PERAMPALADAS, K., DOBLE, B. & PULLENAYEGUM, E. 2014. Comparing EQ-5D valuation studies: a systematic review and methodological reporting checklist. *Medical Decision Making,* 34**,** 8-20.

XIE, F., PULLENAYEGUM, E., GAEBEL, K., BANSBACK, N., BRYAN, S., OHINMAA, A., POISSANT, L. & JOHNSON, J. A. 2016. A time trade-off-derived value set of the EQ-5D-5L for Canada. *Medical Care,* 54**,** 98-105.

YANG, Y., ROWEN, D., BRAZIER, J., TSUCHIYA, A., YOUNG, T. & LONGWORTH, L. 2015. An exploratory study to test the impact on three "bolt-on" items to the EQ-5D. *Value in Health,* 18**,** 52-60.

YANG, Z., LUO, N., BONSEL, G., BUSSCHBACH, J. & STOLK, E. 2017. Selecting health states for EQ-5D-3L valuation studies: statistical considerations matter. *Value in Health,* (Epub).

ZHANG, Y., COELLO, P. A., BROŻEK, J., WIERCIOCH, W., ETXEANDIA-IKOBALTZETA, I., AKL, E. A., MEERPOHL, J. J., ALHAZZANI, W., CARRASCO-LABRA, A. & MORGAN, R. L. 2017. Using patient values and preferences to inform the importance of health outcomes in practice guideline development following the GRADE approach. *Health and Quality of Life Outcomes,* 15**,** 52.

APPENDICES

Appendix 1: EQ-5D-3L Questionnaire

By placing a tick in one box in each group below, please indicate which statements best describe your own health state today.

Mobility

I have no problems in walking about ❑
I have some problems in walking about ❑
I am confined to bed ❑

Self-Care

I have no problems with self-care ❑
I have some problems washing or dressing myself ❑
I am unable to wash or dress myself ❑

Usual Activities (*e.g. work, study, housework, family or leisure activities)*

I have no problems with performing my usual activities ❑
I have some problems with performing my usual activities ❑
I am unable to perform my usual activities ❑

Pain/Discomfort

I have no pain or discomfort ❑
I have moderate pain or discomfort ❑
I have extreme pain or discomfort ❑

Anxiety/Depression

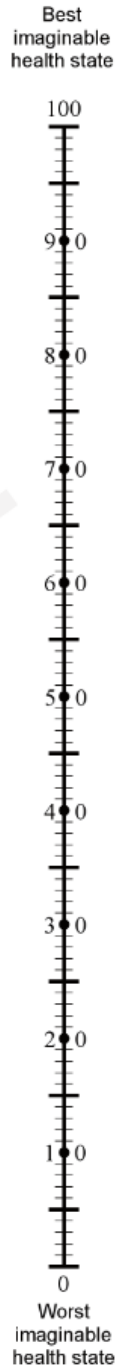
I am not anxious or depressed ❑
I am moderately anxious or depressed ❑
I am extremely anxious or depressed ❑

Source: EQ-5D-3L User Guide

To help people say how good or bad a health state is, we have drawn a scale (rather like a thermometer) on which the best state you can magine is marked **100** and the worst state you can imagine is marked **0**.

We would like you to indicate on this scale how good or bad your own health is today, in your opinion. Please do this by drawing a line from the box below to whichever point on the scale indicates how good or bad your health state is today.

Your own health state today



Source: EQ-5D-3L User Guide

Appendix 2: EQ-5D-5L Questionnaire

Under each heading, please tick the **ONE** box that best describes your health **TODAY**

MOBILITY

I have no problems in walking about ❑
I have slight problems in walking about ❑
I have moderate problems in walking about ❑
I have severe problems in walking about ❑
I am unable to walk about ❑

SELF-CARE

I have no problems washing or dressing myself ❑
I have slight problems washing or dressing myself ❑
I have moderate problems washing or dressing myself ❑
I have severe problems washing or dressing myself ❑
I am unable to wash or dress myself ❑

USUAL ACTIVITIES *(e.g. work, study, housework, family or leisure activities)*

I have no problems doing my usual activities ❑
I have slight problems doing my usual activities ❑
I have moderate problems doing my usual activities ❑
I have severe problems doing my usual activities ❑
I am unable to do my usual activities ❑

PAIN / DISCOMFORT

I have no pain or discomfort ❑
I have slight pain or discomfort ❑
I have moderate pain or discomfort ❑
I have severe pain or discomfort ❑
I have extreme pain or discomfort ❑

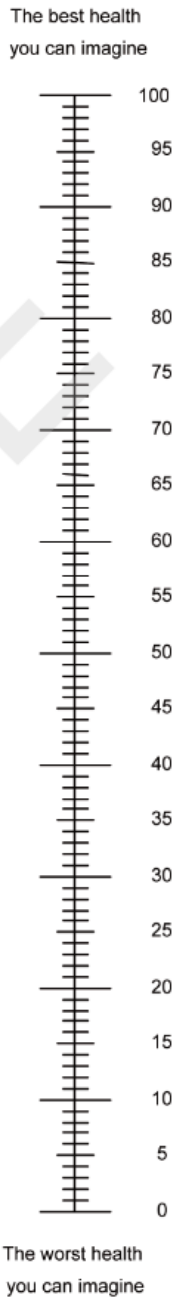
ANXIETY / DEPRESSION

I am not anxious or depressed ❑
I am slightly anxious or depressed ❑
I am moderately anxious or depressed ❑
I am severely anxious or depressed ❑
I am extremely anxious or depressed ❑

Source: EQ-5D-5L User Guide

- We would like to know how good or bad your health is **TODAY**.
- This scale is numbered from **0** to **100**.
- **100** means the best health you can imagine.
  0 means the worst health you can imagine.
- Mark an **X** on the scale to indicate how your health is **TODAY**.
- Now, please write the number you marked on the scale in the box below.

YOUR HEALTH TODAY = ☐



Source: EQ-5D-5L User Guide

Appendix 3: SF-6D Questionnaire

By placing a tick in one box in each group below, please indicate which statements best describe your own health state for the past 4 weeks.

Physical functioning

a. Your health does not limit you in *vigorous activities* ❑
b. Your health limits you a little in *vigorous activities* ❑
c. Your health limits you a little in *moderate activities* ❑
d. Your health limits you a lot in *moderate activities* ❑
e. Your health limits you *a little in bathing and dressing* ❑
f. Your health limits you *a lot in bathing and dressing* ❑

Role limitation

a. You have no problems with your work or other regular daily activities as a result of your physical health or other emotional problems. ❑
b. You are limited in the kind of work or other activities as a result of your physical health. ❑
c. You accomplish less than you would like as a result of emotional problems. ❑
d. You are limited in the kind of work or other activities as a result of your physical health and accomplish less than you would like as a result of emotional problems. ❑

Social functioning

a. Your health limits your social activities *none of the time.* ❑
b. Your health limits your social activities a *little of the time.* ❑
c. Your health limits your social activities *some of the time.* ❑
d. Your health limits your social activities *most of the time.* ❑
e. Your health limits your social activities *all of the time.* ❑

Pain

a. You have *no* pain ❑
b. You have pain but it does not interfere with your normal work (both outside the home and housework). ❑
c. You have pain that interferes with your normal work (both outside the home and housework) *a little bit.* ❑
d. You have pain that interferes with your normal work (both outside the home and housework) *moderately.* ❑
e. You have pain that interferes with your normal work (both outside the home and housework) *quite a bit.* ❑
f. You have pain that interferes with your normal work (both outside the home and housework) *extremely.* ❑

Mental health

a. You feel tense or downhearted and low *none of the time.* ❑
b. You feel tense or downhearted and low *a little of the time.* ❑
c. You feel tense or downhearted and low *some of the time.* ❑
d. You feel tense or downhearted and low *most of the time.* ❑
e. You feel tense or downhearted and low *all of the time.* ❑

Vitality

a. You have a lot of energy *all of the time.* ❑
b. You have a lot of energy *most of the time.* ❑
c. You have a lot of energy *some of the time.* ❑
d. You have a lot of energy *a little of the time.* ❑
e. You have a lot of energy *none of the time.* ❑

Source: www.sheffield.ac.uk

Appendix 4: SF-6D Health States Selected in Study A

Serial Number	Health state	PH	RL	SF	PN	MH	VT
1	111111	1	1	1	1	1	1
2	111112	1	1	1	1	1	2
3	111212	1	1	1	2	1	2
4	111215	1	1	1	2	1	5
5	111222	1	1	1	2	2	2
6	111453	1	1	1	4	5	3
7	111621	1	1	1	6	2	1
8	112111	1	1	2	1	1	1
9	112221	1	1	2	2	2	1
10	112521	1	1	2	5	2	1
11	112543	1	1	2	5	4	3
12	113411	1	1	3	4	1	1
13	114212	1	1	4	2	1	2
14	114244	1	1	4	2	4	4
15	115653	1	1	5	6	5	3
16	121111	1	2	1	1	1	1
17	121112	1	2	1	1	1	2
18	121122	1	2	1	1	2	2
19	121212	1	2	1	2	1	2
20	122112	1	2	2	1	1	2
21	122211	1	2	2	2	1	1
22	122233	1	2	2	2	3	3
23	122425	1	2	2	4	2	5
24	122622	1	2	2	6	2	2
25	122653	1	2	2	6	5	3
26	124114	1	2	4	1	1	4
27	124314	1	2	4	3	1	4
28	131542	1	3	1	5	4	2
29	132425	1	3	2	4	2	5
30	132524	1	3	2	5	2	4
31	133132	1	3	3	1	3	2
32	133511	1	3	3	5	1	1
33	134322	1	3	4	3	2	2
34	134352	1	3	4	3	5	2
35	135332	1	3	5	3	3	2
36	141653	1	4	1	6	5	3

Serial Number	Health state	PH	RL	SF	PN	MH	VT
37	142154	1	4	2	1	5	4
38	142631	1	4	2	6	3	1
39	143611	1	4	3	6	1	1
40	144113	1	4	4	1	1	3
41	144144	1	4	4	1	4	4
42	144241	1	4	4	2	4	1
43	144341	1	4	4	3	4	1
44	145133	1	4	5	1	3	3
45	145353	1	4	5	3	5	3
46	211111	2	1	1	1	1	1
47	211211	2	1	1	2	1	1
48	211212	2	1	1	2	1	2
49	211221	2	1	1	2	2	1
50	212442	2	1	2	4	4	2
51	212453	2	1	2	4	5	3
52	213114	2	1	3	1	1	4
53	213323	2	1	3	3	2	3
54	213345	2	1	3	3	4	5
55	214411	2	1	4	4	1	1
56	214535	2	1	4	5	3	5
57	215154	2	1	5	1	5	4
58	221211	2	2	1	2	1	1
59	221212	2	2	1	2	1	2
60	221432	2	2	1	4	3	2
61	221535	2	2	1	5	3	5
62	222113	2	2	2	1	1	3
63	222121	2	2	2	1	2	1
64	222122	2	2	2	1	2	2
65	222212	2	2	2	2	1	2
66	223451	2	2	3	4	5	1
67	223511	2	2	3	5	1	1
68	224112	2	2	4	1	1	2
69	224223	2	2	4	2	2	3
70	224612	2	2	4	6	1	2
71	232111	2	3	2	1	1	1
72	233551	2	3	3	5	5	1
73	234233	2	3	4	2	3	3
74	234551	2	3	4	5	5	1

Serial Number	Health state	PH	RL	SF	PN	MH	VT
75	235224	2	3	5	2	2	4
76	241531	2	4	1	5	3	1
77	241545	2	4	1	5	4	5
78	241635	2	4	1	6	3	5
79	243432	2	4	3	4	3	2
80	243634	2	4	3	6	3	4
81	244313	2	4	4	3	1	3
82	311222	3	1	1	2	2	2
83	311233	3	1	1	2	3	3
84	311655	3	1	1	6	5	5
85	312255	3	1	2	2	5	5
86	312332	3	1	2	3	3	2
87	312455	3	1	2	4	5	5
88	312552	3	1	2	5	5	2
89	313532	3	1	3	5	3	2
90	314631	3	1	4	6	3	1
91	315515	3	1	5	5	1	5
92	321122	3	2	1	1	2	2
93	321144	3	2	1	1	4	4
94	321221	3	2	1	2	2	1
95	321335	3	2	1	3	3	5
96	321455	3	2	1	4	5	5
97	322134	3	2	2	1	3	4
98	322635	3	2	2	6	3	5
99	322644	3	2	2	6	4	4
100	323135	3	2	3	1	3	5
101	323153	3	2	3	1	5	3
102	323333	3	2	3	3	3	3
103	323431	3	2	3	4	3	1
104	323433	3	2	3	4	3	3
105	323443	3	2	3	4	4	3
106	323632	3	2	3	6	3	2
107	323644	3	2	3	6	4	4
108	323645	3	2	3	6	4	5
109	324125	3	2	4	1	2	5
110	325455	3	2	5	4	5	5
111	331244	3	3	1	2	4	4
112	332113	3	3	2	1	1	3

Serial Number	Health state	PH	RL	SF	PN	MH	VT
113	332145	3	3	2	1	4	5
114	332411	3	3	2	4	1	1
115	333154	3	3	3	1	5	4
116	333225	3	3	3	2	2	5
117	333333	3	3	3	3	3	3
118	333433	3	3	3	4	3	3
119	334254	3	3	4	2	5	4
120	341123	3	4	1	1	2	3
121	342322	3	4	2	3	2	2
122	342353	3	4	2	3	5	3
123	343214	3	4	3	2	1	4
124	343312	3	4	3	3	1	2
125	343325	3	4	3	3	2	5
126	344145	3	4	4	1	4	5
127	344344	3	4	4	3	4	4
128	345122	3	4	5	1	2	2
129	345623	3	4	5	6	2	3
130	411245	4	1	1	2	4	5
131	412152	4	1	2	1	5	2
132	413144	4	1	3	1	4	4
133	413333	4	1	3	3	3	3
134	413414	4	1	3	4	1	4
135	413511	4	1	3	5	1	1
136	414511	4	1	4	5	1	1
137	414522	4	1	4	5	2	2
138	415424	4	1	5	4	2	4
139	421314	4	2	1	3	1	4
140	422655	4	2	2	6	5	5
141	423333	4	2	3	3	3	3
142	423343	4	2	3	3	4	3
143	423433	4	2	3	4	3	3
144	424421	4	2	4	4	2	1
145	424554	4	2	4	5	5	4
146	424643	4	2	4	6	4	3
147	425133	4	2	5	1	3	3
148	425521	4	2	5	5	2	1
149	431144	4	3	1	1	4	4
150	431435	4	3	1	4	3	5

Serial Number	Health state	PH	RL	SF	PN	MH	VT
151	431443	4	3	1	4	4	3
152	431623	4	3	1	6	2	3
153	432255	4	3	2	2	5	5
154	432623	4	3	2	6	2	3
155	433142	4	3	3	1	4	2
156	433333	4	3	3	3	3	3
157	433433	4	3	3	4	3	3
158	433541	4	3	3	5	4	1
159	434654	4	3	4	6	5	4
160	441132	4	4	1	1	3	2
161	442343	4	4	2	3	4	3
162	443144	4	4	3	1	4	4
163	443215	4	4	3	2	1	5
164	443222	4	4	3	2	2	2
165	443335	4	4	3	3	3	5
166	445321	4	4	5	3	2	1
167	511114	5	1	1	1	1	4
168	512242	5	1	2	2	4	2
169	512551	5	1	2	5	5	1
170	513354	5	1	3	3	5	4
171	513531	5	1	3	5	3	1
172	515332	5	1	5	3	3	2
173	521424	5	2	1	4	2	4
174	522321	5	2	2	3	2	1
175	523554	5	2	3	5	5	4
176	523634	5	2	3	6	3	4
177	524442	5	2	4	4	4	2
178	524644	5	2	4	6	4	4
179	525112	5	2	5	1	1	2
180	525311	5	2	5	3	1	1
181	531635	5	3	1	6	3	5
182	532124	5	3	2	1	2	4
183	532455	5	3	2	4	5	5
184	532554	5	3	2	5	5	4
185	533331	5	3	3	3	3	1
186	534133	5	3	4	1	3	3
187	534544	5	3	4	5	4	4
188	534555	5	3	4	5	5	5

Serial Number	Health state	PH	RL	SF	PN	MH	VT
189	534625	5	3	4	6	2	5
190	534644	5	3	4	6	4	4
191	535422	5	3	5	4	2	2
192	535544	5	3	5	5	4	4
193	535545	5	3	5	5	4	5
194	535554	5	3	5	5	5	4
195	535555	5	3	5	5	5	5
196	535645	5	3	5	6	4	5
197	541432	5	4	1	4	3	2
198	541531	5	4	1	5	3	1
199	541622	5	4	1	6	2	2
200	542325	5	4	2	3	2	5
201	542345	5	4	2	3	4	5
202	542524	5	4	2	5	2	4
203	543344	5	4	3	3	4	4
204	543624	5	4	3	6	2	4
205	544223	5	4	4	2	2	3
206	544352	5	4	4	3	5	2
207	544555	5	4	4	5	5	5
208	544633	5	4	4	6	3	3
209	544644	5	4	4	6	4	4
210	544654	5	4	4	6	5	4
211	545122	5	4	5	1	2	2
212	545422	5	4	5	4	2	2
213	545523	5	4	5	5	2	3
214	545622	5	4	5	6	2	2
215	545644	5	4	5	6	4	4
216	545654	5	4	5	6	5	4
217	545655	5	4	5	6	5	5
218	612442	6	1	2	4	4	2
219	613143	6	1	3	1	4	3
220	614321	6	1	4	3	2	1
221	614434	6	1	4	4	3	4
222	615144	6	1	5	1	4	4
223	621221	6	2	1	2	2	1
224	621451	6	2	1	4	5	1
225	622513	6	2	2	5	1	3
226	623133	6	2	3	1	3	3

Serial Number	Health state	PH	RL	SF	PN	MH	VT
227	624115	6	2	4	1	1	5
228	624142	6	2	4	1	4	2
229	624331	6	2	4	3	3	1
230	624343	6	2	4	3	4	3
231	625141	6	2	5	1	4	1
232	625213	6	2	5	2	1	3
233	631231	6	3	1	2	3	1
234	631354	6	3	1	3	5	4
235	631355	6	3	1	3	5	5
236	632121	6	3	2	1	2	1
237	633122	6	3	3	1	2	2
238	634124	6	3	4	1	2	4
239	634545	6	3	4	5	4	5
240	635244	6	3	5	2	4	4
241	635255	6	3	5	2	5	5
242	635544	6	3	5	5	4	4
243	635554	6	3	5	5	5	4
244	641424	6	4	1	4	2	4
245	641622	6	4	1	6	2	2
246	642612	6	4	2	6	1	2
247	644545	6	4	4	5	4	5
248	645555	6	4	5	5	5	5
249	645655	6	4	5	6	5	5

PH Physical functioning, RL Role limitation, SF Social functioning, PN Pain, MH Mental health, VT Vitality.

Appendix 5: EQ-5D-3L Health States Selected in Study A

Serial Number	Health State	MO	SC	UA	PD	AD
1	11111	1	1	1	1	1
2	11112	1	1	1	1	2
3	11113	1	1	1	1	3
4	11121	1	1	1	2	1
5	11122	1	1	1	2	2
6	11131	1	1	1	3	1
7	11133	1	1	1	3	3
8	11211	1	1	2	1	1
9	11312	1	1	3	1	2
10	12111	1	2	1	1	1
11	12121	1	2	1	2	1
12	12211	1	2	2	1	1
13	12222	1	2	2	2	2
14	12223	1	2	2	2	3
15	13212	1	3	2	1	2
16	13311	1	3	3	1	1
17	13332	1	3	3	3	2
18	21111	2	1	1	1	1
19	21133	2	1	1	3	3
20	21222	2	1	2	2	2
21	21232	2	1	2	3	2
22	21312	2	1	3	1	2
23	21323	2	1	3	2	3
24	22112	2	2	1	1	2
25	22121	2	2	1	2	1
26	22122	2	2	1	2	2
27	22222	2	2	2	2	2
28	22233	2	2	2	3	3
29	22323	2	2	3	2	3
30	22331	2	2	3	3	1
31	23232	2	3	2	3	2
32	23313	2	3	3	1	3
33	23321	2	3	3	2	1
34	32211	3	2	2	1	1
35	32223	3	2	2	2	3
36	32232	3	2	2	3	2

Serial Number	Health State	MO	SC	UA	PD	AD
37	32313	3	2	3	1	3
38	32331	3	2	3	3	1
39	33212	3	3	2	1	2
40	33232	3	3	2	3	2
41	33321	3	3	3	2	1
42	33323	3	3	3	2	3
43	33333	3	3	3	3	3

MO Mobility, SC Self-care, UA Usual activities, PD Pain/Discomfort, AD Anxiety/Depression.

Appendix 6: EQ-5D-5L Health States Selected in Study C

Block	Health State	MO	SC	UA	PD	AD
1	11221	1	1	2	2	1
	11235	1	1	2	3	5
	54231	5	4	2	3	1
	51451	5	1	4	5	1
	34515	3	4	5	1	5
	35245	3	5	2	4	5
	12514	1	2	5	1	4
	45144	4	5	1	4	4
	12111	1	2	1	1	1
	55555	5	5	5	5	5
2	12543	1	2	5	4	3
	12121	1	2	1	2	1
	43542	4	3	5	4	2
	34155	3	4	1	5	5
	52215	5	2	2	1	5
	45133	4	5	1	3	3
	32443	3	2	4	4	3
	23514	2	3	5	1	4
	11211	1	1	2	1	1
	55555	5	5	5	5	5
3	45233	4	5	2	3	3
	55233	5	5	2	3	3
	31525	3	1	5	2	5
	52455	5	2	4	5	5
	12244	1	2	2	4	4
	13313	1	3	3	1	3
	25122	2	5	1	2	2
	11421	1	1	4	2	1
	21111	2	1	1	1	1
	55555	5	5	5	5	5
4	21112	2	1	1	1	2
	14554	1	4	5	5	4
	12513	1	2	5	1	3

Block	Health State	MO	SC	UA	PD	AD
	44345	4	4	3	4	5
	12344	1	2	3	4	4
	53221	5	3	2	2	1
	54342	5	4	3	4	2
	44125	4	4	1	2	5
	11121	1	1	1	2	1
	55555	5	5	5	5	5
5	43315	4	3	3	1	5
	54153	5	4	1	5	3
	52431	5	2	4	3	1
	24443	2	4	4	4	3
	14113	1	4	1	1	3
	31524	3	1	5	2	4
	15151	1	5	1	5	1
	21315	2	1	3	1	5
	11112	1	1	1	1	2
	55555	5	5	5	5	5
6	12112	1	2	1	1	2
	11212	1	1	2	1	2
	44553	4	4	5	5	3
	21345	2	1	3	4	5
	34244	3	4	2	4	4
	23152	2	3	1	5	2
	43514	4	3	5	1	4
	55424	5	5	4	2	4
	21111	2	1	1	1	1
	55555	5	5	5	5	5
7	13122	1	3	1	2	2
	24553	2	4	5	5	3
	51152	5	1	1	5	2
	11425	1	1	4	2	5
	22434	2	2	4	3	4
	42115	4	2	1	1	5
	35332	3	5	3	3	2
	45413	4	5	4	1	3

Block	Health State	MO	SC	UA	PD	AD
	11211	1	1	2	1	1
	55555	5	5	5	5	5
8	33253	3	3	2	5	3
	23242	2	3	2	4	2
	24342	2	4	3	4	2
	32314	3	2	3	1	4
	12334	1	2	3	3	4
	21334	2	1	3	3	4
	55225	5	5	2	2	5
	53412	5	3	4	1	2
	11112	1	1	1	1	2
	55555	5	5	5	5	5
9	11414	1	1	4	1	4
	25331	2	5	3	3	1
	25222	2	5	2	2	2
	21444	2	1	4	4	4
	31514	3	1	5	1	4
	53243	5	3	2	4	3
	53244	5	3	2	4	4
	35143	3	5	1	4	3
	11121	1	1	1	2	1
	55555	5	5	5	5	5
10	11122	1	1	1	2	2
	52335	5	2	3	3	5
	35311	3	5	3	1	1
	43555	4	3	5	5	5
	24445	2	4	4	4	5
	13224	1	3	2	2	4
	34232	3	4	2	3	2
	42321	4	2	3	2	1
	12111	1	2	1	1	1
	55555	5	5	5	5	5

MO mobility, SC self-care, UA usual activities, PD pain/discomfort, AD anxiety/depression.

Appendix 7: Post-hoc Analysis of Comparing Valuation Score of Mild and Moderate Health States between the Heart Disease and Control Groups in Study A.

EQ-5D-3L health states#	**Control (N = 651)**	**Heart diseases (N = 44)**
Mild health states		
Mean (SD)	81.3 (16.7)	89.1 (4.5)
Adjusted mean difference (95% CI)	12.8 (8.0, 17.7) ***	
Moderate health states		
Mean (SD)	68.8 (19.2)	64.9 (15.1)
Adjusted mean difference (95% CI)	0.1 (-5.0, 5.2)	

Health states with only one domain at severity level 2 and remaining all the domains at severity level 1 are classified as 'mild,' health states with at least one domain at severity level 3 are classified as 'severe,' and remaining health states are classified as 'moderate.'

Adjusted mean difference: mean scores of the heart disease patient group minus mean scores of the control group, estimated by ordinary least-square regression model, with adjustment for health state descriptors, disutility due to severe problems, and sociodemographic covariates (see Methods section).

*** p <0.001.

ORIGINAL PUBLICATIONS

PUBLICATION I

How is the most severe health state being valued by the general population?

Mihir Gandhi, Julian Thumboo, Hwee Lin Wee, Nan Luo, Yin Bun Cheung

Health and Quality of Life Outcomes 2014, 12:161
https://doi.org/10.1186/s12955-014-0161-9

Gandhi et al. Health and Quality of Life Outcomes 2014, 12:161
http://www.hqlo.com/content/12/1/161

HEALTH AND QUALITY OF LIFE OUTCOMES

RESEARCH Open Access

How is the most severe health state being valued by the general population?

Mihir Gandhi[1,2,3*], Julian Thumboo[4,5], Hwee-Lin Wee[6], Nan Luo[7] and Yin-Bun Cheung[1,2,3]

Abstract

Background: It has been reported that valuation of health states that are close to death, such as the most severe health state, can be affected by health state valuation procedure, and their utility values are difficult to predict. We examined how the most severe health states of Short Form-6 dimension (SF-6D) and EuroQoL-5 dimension-3 level (EQ-5D-3L) were valued by the Singapore general population.

Methods: Overall, 249 SF-6D and 42 EQ-5D-3L states were valued by two separate samples from the Singapore general population using the visual analogue scale (VAS) method. Ordinary least-square regression model was employed to explain deficit in the valuation of the most severe state using the health state descriptors.

Results: A total of 1021 participants from the SF-6D sample and 1015 participants from the EQ-5D-3L sample were included in the analysis. We observed that 67% of the SF-6D participants and 74% of the EQ-5D-3L participants considered the most severe state worse than dead. The most severe state had mean VAS valuation scores more than 20–25 points lower than the adjacent states that are better by only one level in only one dimension. SF-6D VAS valuation score for the most severe state was 27 points and 12 points lower than expected according to the health state descriptors among the participants who considered the most severe state worse than dead and better than dead, respectively. Similar results were found for the EQ-5D-3L valuation.

Conclusions: The most severe health state was valued lower than expected according to its descriptors.

Keywords: EQ-5D, SF-6D, Utility, Visual analogue scale, Worse than dead

Background

There is an increasing demand on evaluating the outcome of health-care interventions using cost utility analysis (CUA), and various health regulatory bodies consider it the main approach for evaluating the outcome of health-care interventions [1]. CUA involves quality adjusted life years (QALYs), which are estimated as the time spent in a health state multiplied by its utility. QALYs are very useful as they capture changes in both quality and quantity of life. Health states from generic health outcome instruments such as EuroQoL-5 dimension (EQ-5D), Short Form-6 dimension (SF-6D) and Health Utility Index Mark 3 (HUI3) are valued by the general population using one or more valuation methods such as time trade-off (TTO), standard gamble (SG) and visual analogue scale (VAS) [2]. A number of health states with a mixture of severity levels, including perfect health, dead and the most severe health state described by the instrument, are valued by each participant. The utility of perfect health is valued 1 and dead is valued 0. A negative utility value represents that the health state is considered 'worse than dead'.

The most severe health state is an important health state in valuation studies. Usually either the most severe health state or dead state is the least valued state and hence decides the lower bound of utility values [3]. For example, the valuation studies in several European countries asked participants to value the most severe health state along with the perfect health and dead state twice because these states are anchoring states (lower or upper bound of the utility values) [4]. Some valuation methods, such as chained TTO and chained SG, also use the most severe health state as the 'temporary health state' and ask participants to trade-off or gamble the other health

* Correspondence: mihir.gandhi@scri.edu.sg
[1]Centre for Quantitative Medicine, Office of Clinical Sciences, Duke-NUS Graduate Medical School, Singapore, Singapore
[2]Biostatistics, Singapore Clinical Research Institute, Singapore, Singapore
Full list of author information is available at the end of the article

states between the perfect health and the most severe health state [5].

The most severe health state is often labelled as 'all-worst' or 'pits' [6-10]. There is a general consensus that adding a disease label in disease-specific health state descriptions can have an impact on health state value, possibly due to prior knowledge or preconception of the disease [11]. However, no study to our knowledge has investigated the impact of labeling a generic health state. Valuation of health states based on generic instruments is the most common method for eliciting the population preferences in CUA, thus how the most severe health state (in the presence of a label) is valued needs attention.

Drawing on data from a valuation study of the two most commonly used generic instruments, namely SF-6D and EQ-5D-3 level (EQ-5D-3L), we aim to examine how the most severe health state is valued by the participants. In this study, the most severe health state was labelled as 'all-worst' state. A valuation study in multi-ethnic Asian general population showed that the majority of participants felt that 'all-worst' is a better description than 'pits' for the most severe health state of EQ-5D [12].

Methods

Valuation survey procedures

A cross-sectional, face-to-face survey of health state valuation for SF-6D and EQ-5D-3L using the VAS method was conducted in 2009 from a representative sample of the general population of Singapore, a multi-ethnic Asian country. A multi-stage sampling approach was used to randomly select residential blocks, within which households were selected. Potential participants who satisfied the pre-set recruitment quotas for ethnicity (400 Chinese, 400 Malay, 234 Indian), gender (50% Female) and age (30% for 21–34 years, 40% for 35–49 years, 30% for 50+ years) were interviewed. Within each race, there was a quota that half of the participants would use English and the remaining half would use their native language for the interviews, i.e. Mandarin for Chinese, Malay for Malays and Tamil for Indians.

Two separate samples of 1034 participants each were selected for the SF-6D and EQ-5D-3L health states valuation. The SF-6D consists of 6 dimensions (physical functioning, role limitations, social functioning, pain, mental health and vitality) and each dimension has 4 to 6 levels. Thus, SF-6D describes a total of 18,000 health states. A subset of 249 SF-6D states was selected (out of 18,000) for the valuation based on the protocol of Brazier *et al*. [6]. The most severe state ('645655' for SF-6D) was labeled as 'all-worst'. Each participant was asked to compare between 'dying now' and 'living for the rest of his/her life in all-worst', from which the less desirable state was assigned a value 0 on the VAS. Each participant was then asked to value a unique set of 6 states from the subset of SF-6D states and either dead or the 'all-worst' state, depending on which one was not valued earlier at 0. The unique set of 6 health states were assigned to each participant in a way that they spread widely over the valuation space. A 100-point "feeling thermometer" with endpoints of 100 (most desirable, i.e. perfect health) and 0 (least desirable) was used as the VAS. The participants were required to indicate where they would rate each of the assigned states on the "feeling thermometer" by imagine themselves in that state for the rest of their life without changing. The participants were allowed to value more than one health state at the same level of VAS.

The valuation of EQ-5D-3L states was carried out in a similar way as SF-6D. The EQ-5D-3L consists of 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) with 3 levels each (no problems, some problems, and extreme problems) and thus describes 243 health states. A subset of 42 EQ-5D-3L states was selected based on the protocol of Dolan [13]. The most severe state ('33333') was labeled as 'all-worst'. Unconscious state was also valued in addition to the other 6 assigned states.

The study was approved by SingHealth Centralized Institutional Review Board.

Analyses

Participants who met the following criteria were excluded from our analysis: a) valued less than 3 health states, b) did not value dead or the 'all-worst' state, c) valued dead or the 'all-worst' state or unconscious state higher than all the other states, d) gave the same valuation score to all the health states, e) self-reported or rated by the interviewers as having poor understanding of health states description or valuation tasks. The valuation score used in the analyses was 'raw' VAS valuation score ranging from 0 (worst possible score) to 100 (best possible score). We did not transform the valuation scores to utility to avoid an impact of the transformation on the estimated valuation score [3].

Mean valuation scores of the health states are presented using line graphs for a subset of selected health states. As the study has valued many health states, to maintain visual clarity we did not include all the valued health states. The lowest ten health states with the least valuation scores near the dead state were included in the graphs. Health states with higher valuation scores were systematically skipped for the graphical presentation. The valuation scores among the participants who considered the 'all-worst' state worse than dead and who considered the 'all-worst' state better than dead were presented separately in the graphs.

We performed ordinary least-square (OLS) regression model for EQ-5D-3L with valuation score as the dependent variable and indicator variables representing level of severity for each dimension as the independent variables, with an intercept [4]. That is, included 2 indicator variables for each of the 5 dimensions of EQ-5D-3L. We also included an indicator variable (N_3) to take into account of additional disutilities when severe problem (level 3) is reported on at least one dimension. In addition to the above commonly used variables, we included two indicator variables D_1 and D_2 and their interaction: D_1 represented the participant who considered the 'all-worst' state worse than dead and D_2 represented the 'all-worst' state. This model helped to assess whether there was a deficit in the valuation score for the most severe state even after taking the descriptors (levels and dimensions) into account. It also assessed possible impact of considering the most severe state worse than death on its valuation.

Similar regression analysis model was used to study SF-6D valuation scores. For the variable N_3, the severe level was defined as levels 4–6 for physical functioning, levels 3–4 for role limitation, level 4–5 for social functioning, mental health and vitality, and level 5–6 for pain [6].

All the health states valued in the study were included in the regression analysis. Perfect health state was not included in the models, as it was assigned a value 100 on VAS. The dead and unconscious states were also excluded from the regression analyses as they do not represent any health states/dimensions of SF-6D or EQ-5D-3L. Since each of the participants valued 6 health states, we used the Eicker-Huber-White robust standard error for clustered data for statistical inference [14]. A P-value less than 0.05 was considered statistically significant. All the analyses were carried out using Stata/MP 10.1 for Windows.

Results

Demographic and health characteristics information was received from all 1034 participants of the SF-6D sample. Seven participants valued dead higher than all the other states; 1 participant valued the 'all-worst' state higher than all the other states; and 5 participants were observed to have poor understanding of heath states description and/or valuation tasks. Hence, these 13 participants were excluded from the SF-6D related analysis. Table 1 shows demographic and health characteristics of 1021 participants for the SF-6D valuation that were included in the analysis. Due to the pre-specified quota for gender, age and ethnicity, the demographic characteristics of enrolled participants were similar to what was planned. The majority of the SF-6D participants were married (n = 765, 75%), employed/self-employed (n = 659, 65%), had at least secondary education (n = 844, 83%), and self-reported good to excellent general health (n = 947, 93%).

Table 1 Demographic and health characteristics of the study sample participants

Characteristics, n (%)	Participants completing the SF-6D (N = 1021)	Participants completing the EQ-5D-3L (N = 1015)
Female	521 (51.0)	512 (50.4)
Age (years)		
21-29	190 (18.6)	194 (19.1)
30-39	222 (21.7)	228 (22.5)
40-49	272 (26.6)	269 (26.5)
50-59	206 (20.2)	200 (19.7)
60+	131 (12.8)	124 (12.2)
Ethnicity		
Chinese	392 (38.4)	387 (38.1)
Malay	396 (38.8)	399 (39.3)
Indian	233 (22.8)	229 (22.6)
Education level		
Primary (6 years) or less	177 (17.3)	190 (18.7)
Secondary (11 years)	562 (55.0)	576 (56.8)
Diploma/degree or higher	282 (27.6)	249 (24.5)
Married/partner	765 (74.9)	761 (75.0)
Employed or self-employed	659 (64.5)	643 (63.4)
General health status		
Poor	4 (0.4)	8 (0.8)
Fair	70 (6.9)	99 (9.8)
Good	443 (43.4)	424 (41.8)
Very good	419 (41.0)	383 (37.7)
Excellent	85 (8.3)	101 (10.0)
Self-reported health on EQ-5D-3L VAS, Mean (SD)	84.5 (11.2)	83.0 (12.1)

VAS = visual analogue scale.

Similarly in the EQ-5D-3L sample, 12 participants valued dead higher than all the other states; 2 participants valued the unconscious state higher than all the other states; 1 participant did not value the 'all-worst' state; and 4 participants were observed to have poor understanding of heath states description and/or valuation tasks. Hence, 19 participants were excluded from the EQ-5D-3L related analysis. The demographic and health characteristics of the EQ-5D-3L participants were similar to the SF-6D participants (Table 1).

In the SF-6D valuation, except the 'all-worst' state, no other health state was valued worse than dead state. The majority of participants (n = 681, 67%) considered the 'all-worst' state worse than dead. The mean valuation score given to the 'all-worst' state was 12.8 (SD = 14.0) among the participants who considered the 'all-worst'

state better than dead. On the other hand, dead state was valued with mean valuation score of 11.2 (SD = 12.5) among the participants who considered the 'all-worst' state worse than dead. The mean valuation scores for selected SF-6D health states are shown in Figure 1. For the participants who considered the 'all-worst' state worse than dead, there was a difference of more than 30 points in the mean valuation score between the 'all-worst' state ('645655') and its adjacent health states that are only one level different in one dimension ('545655' and '645555'). For the participants who considered the 'all-worst' state better than dead the corresponding difference ranged from 9 to 17 points.

Similar to the SF-6D results, only the 'all-worst' state was valued worse than dead, and the majority of participants (n = 753, 74%) considered the 'all-worst' state worse than dead in the EQ-5D-3L valuation. The mean valuation score of the 'all-worst' state was 11.4 (SD = 12.3) among the participants who considered the 'all-worst' state better than dead; and the mean valuation score of dead was 15.2 (SD = 19.7) among the participants who considered the 'all-worst' state worse than dead. Similar to Figures 1, 2 also shows a difference of 25 points in the mean valuation score between the 'all-worst' state ('33333') and its adjacent state '33323' for the participants who considered the 'all-worst' worse than dead. For the participants who considered the 'all-worst' better than dead the corresponding difference was 11 points.

The regression intercept showed that, among the participants who considered the 'all-worst' state better than dead in the SF-6D sample, the estimated valuation score was 80.2 (95% CI: 77.3, 83.1) when all the dimensions were at its best (level 1) (Table 2). The coefficient D_1 showed that the participants who considered the 'all-worst' state worse than dead scored the other health states higher by 2 points (95% CI: 0.2, 3.8) compared to the participants who considered the 'all-worst' state better than dead. The participants who considered the 'all-worst' state better than dead scored the 'all-worst' state 12.3 points lower (95% CI: −15.8, −8.8; P-value < 0.001) than expected by its descriptors. Furthermore, the participants who considered the 'all-worst' state worse dead scored the 'all-worst' state 27.0 points lower (coefficient: −12.3-14.7 = 27.0; 95% CI: −30.2, −24.0; P-value < 0.001) than expected.

For the EQ-5D-3L valuation, the participants who considered the 'all-worst' state better than dead scored the 'all-worst' state 2.1 points lower (95% CI: −5.4, −1.1; P-value = 0.201) than expected by its descriptors. The participants who considered the 'all-worst' state worse than dead scored the 'all-worst' state 15.8 points lower (coefficient = −2.1-13.7 = −15.8; 95% CI: −18.8, −12.7; P-value < 0.001) than expected (Table 3).

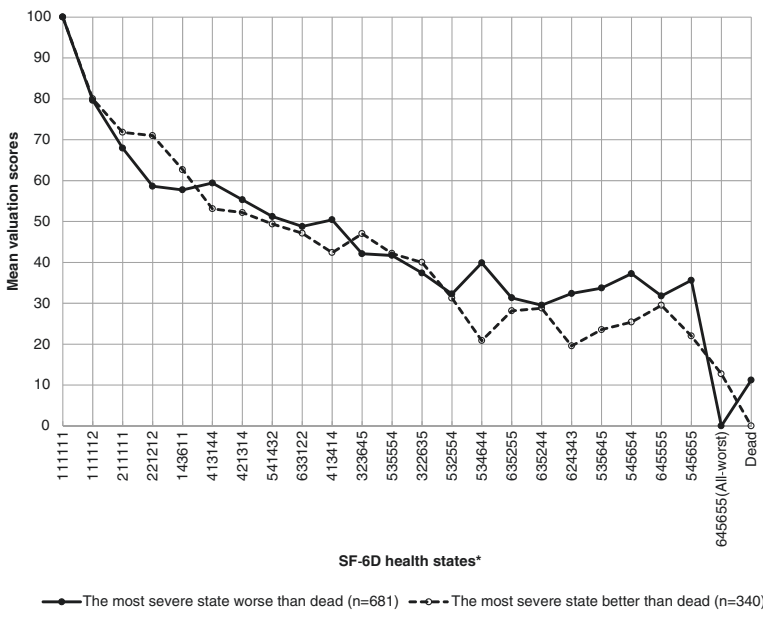


Figure 1 Mean valuation scores of SF-6D health states. *The lowest ten health states with the least valuation scores near the dead state are included in the graphs. Health states with higher valuation scores were systematically skipped for the graphical presentation.

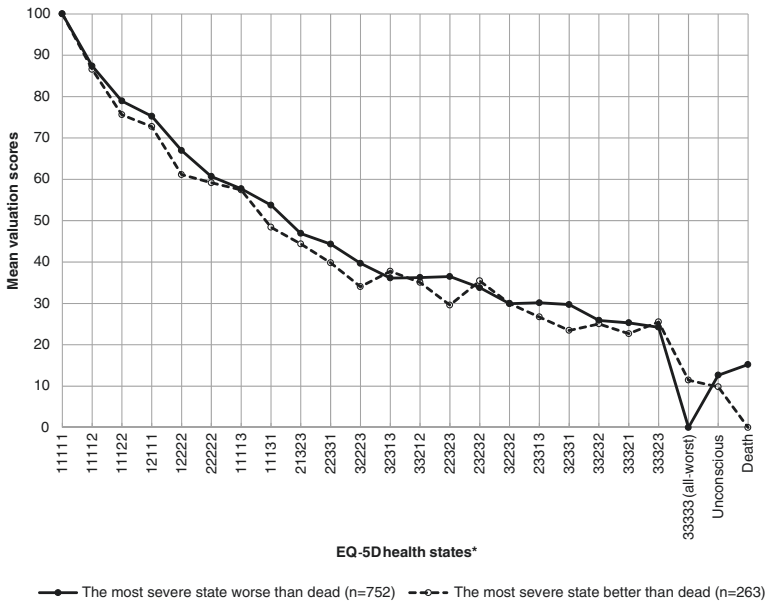


Figure 2 Mean valuation scores of EQ-5D-3L health states. *The lowest ten health states with the least valuation scores near the dead state are included in the graphs. Health states with higher valuation scores were systematically skipped for the graphical presentation.

All the coefficients of indicator variables for different levels of SF-6D and EQ-5D-3L dimensions in the respective regression models were negative and most of them were statistically significant (Tables 2 and 3).

Discussion

We examined the valuation of the most severe state (labeled as 'all-worst') of two of the most widely used generic instruments, SF-6D and EQ-5D-3L, using the VAS method in a general population. We examined the valuation score in two groups of participants - who considered the most severe state worse than dead and who considered it better than dead. We observed a deficit of 20–25 points in the valuation score of the most severe state than the adjacent states that are better by only one level in only one dimension. The regression analysis showed that this deficit could not be explained by the differences in the health state descriptors. In the SF-6D valuation, the unexplained deficit was about 27 points and 12.3 points, respectively, among participants who considered the most severe state worse and better than dead. This is practically significant, as 3.3 points was considered minimally important difference [15,16]. In the valuation of the EQ-5D-3L, the participants who considered the most severe state worse than dead also scored the most severe state 15.8 points lower than expected according to its descriptor, which is bigger than the minimally important difference of 7–8 point [16,17]. The regression models were developed based on the comparative review and user guide for EQ-5D value sets [4]. We used raw VAS valuation score in the analysis without any transformations or rescaling to reduce artifact effect on the regression coefficients [3].

We found a deficit in value in the health state labelled as 'all-worst'. Participants may have valued the most severe state based on the 'all-worst' label rather than the objective description of it [18]. That is, the valuation might be affected by prior belief associated with a worst health condition or an emotional response to the hearing of 'all-worst'. Several studies of disease-specific health states valuation have found that an inclusion of disease label lowered the health state values [11]. For example, a study found that using a label 'breast cancer' reduced health state values [19]. Another study found that the use of mental health labels such as mental handicap, schizophrenia and dementia was associated with lower health state values [20]. The present study showed a similar labeling effect in a generic health state and in a community context. Thus, we suggest to avoid labeling health states in valuation studies.

Furthermore, the difference in the degree of deficit in the values of the most severe state between participants who considered the state worse than dead and those who considered it better than dead might have been

Table 2 Summary of ordinary least-square regression on valuation score of SF-6D health states

Regressor*	Coefficient	95% confidence interval	P-value
The most severe state worse than dead (D_1)	2.0	[0.2, 3.8]	0.029
The most severe state (D_2)	−12.3	[−15.8, −8.8]	<0.001
Interaction of D_1 and D_2	−14.7	[−16.6, −12.9]	<0.001
At least one severe level (N_3)	−2.3	[−4.3, −0.4]	0.018
Physical functioning level 2	−7.5	[−16.6, −12.9]	<0.001
Physical functioning level 3	−7.9	[−9.7, −6.0]	<0.001
Physical functioning level 4	−13.8	[−15.8, −11.7]	<0.001
Physical functioning level 5	−13.9	[−15.9, −11.9]	<0.001
Physical functioning level 6	−22.1	[−24.4, −19.9]	<0.001
Role limitations level 2	−3.7	[−5.2, −2.2]	<0.001
Role limitations level 3	−3.1	[−4.6, −1.6]	<0.001
Role limitations level 4	−2.7	[−4.3, −1.0]	0.002
Social functioning level 2	−3.1	[−4.6, −1.5]	<0.001
Social functioning level 3	−2.1	[−3.8, −0.4]	0.015
Social functioning level 4	−2.9	[−4.8, −1.1]	0.002
Social functioning level 5	−3.2	[−5.1, −1.4]	0.001
Pain level 2	−3.9	[−5.7, −2.2]	<0.001
Pain level 3	−6.3	[−8.1, −4.5]	<0.001
Pain level 4	−6.8	[−8.6, −4.9]	<0.001
Pain level 5	−7.5	[−9.2, −5.7]	<0.001
Pain level 6	−11.1	[−13.0, −9.2]	<0.001
Mental health level 2	−2.1	[−3.8, −0.4]	0.016
Mental health level 3	−2.8	[−4.7, −1.0]	0.002
Mental health level 4	−4.0	[−5.9, −2.1]	<0.001
Mental health level 5	−3.0	[−4.9, −1.1]	0.002
Vitality level 2	−0.3	[−2.1, 1.6]	0.785
Vitality level 3	−5.0	[−6.9, −3.2]	<0.001
Vitality level 4	−5.6	[−7.6, −3.6]	<0.001
Vitality level 5	−10.6	[−12.6, −8.6]	<0.001
Intercept	80.2	[77.3, 83.1]	<0.001
$R^2 = 0.519$			

*Adjusted for all levels of 6 dimensions of SF-6D.

Table 3 Summary of ordinary least-square regression on valuation score of EQ-5D-3L health states

Regressor*	Coefficient	95% confidence interval	P-value
The most severe state worse than dead (D_1)	2.2	[−0.0, 4.5]	0.054
The most severe state (D_2)	−2.1	[−5.4, 1.1]	0.201
Interaction of D_1 and D_2	−13.7	[−15.8, −11.6]	<0.001
At least one severe level (N_3)	−14.8	[−17.4, −12.1]	<0.001
Mobility level 2	−8.2	[−9.4, −6.9]	<0.001
Mobility level 3	−15.0	[−17.0, −13.0]	<0.001
Self-care level 2	−5.1	[−6.5, −3.7]	<0.001
Self-care level 3	−13.9	[−15.7, −12.2]	<0.001
Usual activities level 2	−1.6	[−3.5, 0.3]	0.102
Usual activities level 3	−4.2	[−6.3, −2.0]	<0.002
Pain/discomfort level 2	−5.1	[−6.3, −3.9]	<0.001
Pain/discomfort level 3	−12.4	[−13.6, −11.1]	<0.001
Anxiety/depression level 2	−0.3	[−1.8, 1.2]	0.674
Anxiety/depression level 3	−6.1	[−7.9, −4.2]	<0.001
Intercept	79.9	[77.5, 82.3]	<0.001
$R^2 = 0.642$			

*Adjusted for all levels of 5 dimensions.

affected by end-aversion bias [21]. Some participants might be reluctant to value health states at the extreme end of the VAS scale or the portion of scale near the end. For the participants who consider the most severe health state worse than dead, the valuation procedure involved putting the card that represented the state at the lower end of the VAS scale (score 0). Then, the remaining health states were valued between the two ends of the VAS scale. Thus, the other health states near the lower end of the VAS scale might have been valued higher than the most severe health state as a result of end-aversion bias. On the other hand, for the participants who considered the most severe health state better than dead, the dead state was anchored at 0 and the most severe health state and the other health states were valued similarly between the two ends of the VAS scale.

Our study showed that there was larger unexplained deficit in the valuation of the most severe state in SF-6D than EQ-5D-3L. It was likely due to differences in their descriptive systems. It has been shown by Brazier *et al.* [22] that severe SF-6D states are less severe than the severe EQ-5D-3L states. That is, the participants consider the most severe state of EQ-5D-3L almost like the worst state based on its descriptors and hence leave less room for the other factors which can intensify its severity, whereas in SF-6D there is more scope for other factors, such as the labeling effect.

It should be noted that we valued the health states using the VAS method and hence the study findings are applicable to this valuation method only. Further research would be needed to examine if the present findings are generalizable to other valuation methods like TTO and SG.

Conclusions

We found that the most severe state was valued significantly lower than expected according to its descriptors. The magnitude of the deficit depended on the valuation instrument and whether the respondents considered the most severe state worse or better than dead.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
JT conceived the study, and participated in the study design and coordination. HLW and LN participated in the study design and coordination. YB participated in the design of the study, statistical analysis and interpretation of the results. MG carried out the data analysis and interpretation, and wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements
The study was funded by a programme grant (03/1/27/18/226) from the Biomedical Research Council of Singapore. The last author (YBC) was supported by Singapore Ministry of Health's National Medical Research Council under its Clinician Scientist Award. The authors appreciate the support of Duke-NUS/SingHealth Academic Research Institute and the medical editing assistance of Taara Madhavan (Associate, Duke-NUS Graduate Medical School).

Author details
[1]Centre for Quantitative Medicine, Office of Clinical Sciences, Duke-NUS Graduate Medical School, Singapore, Singapore. [2]Biostatistics, Singapore Clinical Research Institute, Singapore, Singapore. [3]Department of International Health, School of Medicine, University of Tampere, Tampere, Finland. [4]Department of Rheumatology and Immunology, Singapore General Hospital, Singapore, Singapore. [5]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [6]Department of Pharmacy, Faculty of Science, National University of Singapore, Singapore, Singapore. [7]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore.

Received: 6 February 2014 Accepted: 9 October 2014
Published online: 25 October 2014

References

1. 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 Health* 2009, **12**(8):119–200.
2. Green C, Brazier J, Deverill M: **Valuing health-related quality of life. A review of health state valuation techniques.** *Pharmacoeconomics* 2000, **17:**151–165.
3. Lamers L: **The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value sets.** *Med Care* 2007, **45**(3):238–244.
4. Szende A, Oppe M, Devlin N: *EQ-5D Value Sets: Inventory, Comparative Review and User Guide (EuroQol group monographs, Vol. 2)*. Springer; 2007.
5. Wright DR, Wittenberg E, Swan JS, Miksad RA, Prosser LA: **Methods for measuring temporary health States for cost-utility analyses.** *Pharmacoeconomics* 2009, **27**(9):713–723.
6. Brazier J, Roberts J, Deverill M: **The estimation of a preference-based measure of health from the SF-36.** *J Health Econ* 2002, **21**(2):271–292.
7. Raat H, Bonsel GJ, Hoogeveen WC, Essink-Bot ML, Dutch HUI Group: **Feasibility and reliability of a mailed questionnaire to obtain visual analogue scale valuations for health states defined by the Health Utilities Index Mark 3.** *Med Care* 2004, **42**(1):13–18.
8. Luo N, Wang Q, Feeny D, Chen G, Li SC, Thumboo J: **Measuring health preferences for Health Utilities Index Mark 3 health states: a study of feasibility and preference differences among ethnic groups in Singapore.** *Med Decis Making* 2007, **27**(1):61–70.
9. Lam CL, Brazier J, McGhee SM: **Valuation of the SF-6D health states is feasible, acceptable, reliable, and valid in a Chinese population.** *Value Health* 2008, **11**(2):295–303.
10. Kharroubi SA, Brazier JE, McGhee S: **Modeling SF-6D Hong Kong standard gamble health state preference data using a nonparametric Bayesian method.** *Value Health* 2013, **16**(6):1032–1045.
11. Brazier JE, Rowen D, Mavranezouli I, Tsuchiya A, Young T, Yang Y, Barkham M, Ibbotson R: **Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome).** *Health Technol Assess* 2012, **16**(32):1–114.
12. Wee HL, Li SC, Xie F, Zhang XH, Luo N, Cheung YB, Machin D, Fong KY, Thumboo J: **Are Asians comfortable with discussing death in health valuation studies? A study in multi-ethnic Singapore.** *Health Qual Life Outcomes* 2006, **4:**93.
13. Dolan P: **Modeling valuations for EuroQol health states.** *Med Care* 1997, **35**(11):1095–1108.
14. Williams R: **A note on robust variance estimation for cluster-correlated data.** *Biometrics* 2000, **56:**645–646.
15. Walters S, Brazier J: **What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D.** *Health Qual Life Outcomes* 2003, **11:**1–4.
16. Wee H, Machin D, Loke W, Li S, Cheung Y, Luo N, Feeny D, Fong K, Thumboo J: **Assessing differences in utility scores: a comparison of four widely used preference-based instruments.** *Value Health* 2007, **10**(4):256–265.
17. Luo N, Johnson J, Coons S: **Using instrument-defined health state transitions to estimate minimally important differences for four preference-based health-related quality of life instruments.** *Med Care* 2010, **48**(4):365–371.
18. Rowen D, Brazier J, Tsuchiya A, Young T, Ibbotson R: **It's all in the name, or is it? The impact of labeling on health state values.** *Med Decis Making* 2012, **32**(1):31–40.
19. Robinson S, Bryan S: *'Naming and framing': an investigation of the effect of disease labels on health state valuations.* Oxford: Health Economics Study Group Meeting, University of Oxford; 2001.
20. Rabin R, Rosser RM, Butler C: **Impact of diagnosis on utilities assigned to states of illness.** *J R Soc Med* 1993, **86:**444–448.
21. Torrance GW, Feeny D, Furlong W: **Visual analog scales: do they have a role in the measurement of preferences for health states?** *Med Decis Making* 2001, **21**(4):329–334.
22. Brazier J, Roberts J, Tsuchiya A, Busschbach J: **A comparison of the EQ-5D and SF-6D across seven patient groups.** *Health Econ* 2004, **13**(9):873–884.

doi:10.1186/s12955-014-0161-9
Cite this article as: Gandhi *et al.*: How is the most severe health state being valued by the general population? *Health and Quality of Life Outcomes* 2014 **12**:161.

PUBLICATION II

Do chronic disease patients value generic health states differently from individuals with no chronic disease? A case of a multicultural Asian population

Mihir Gandhi, Julian Thumboo, Nan Luo, Hwee Lin Wee, Yin Bun Cheung

Health and Quality of Life Outcomes 2015, 13:8
https://doi.org/10.1186/s12955-014-0200-6

Gandhi et al. Health and Quality of Life Outcomes (2015) 13:8
DOI 10.1186/s12955-014-0200-6

HEALTH AND QUALITY OF LIFE OUTCOMES

RESEARCH Open Access

Do chronic disease patients value generic health states differently from individuals with no chronic disease? A case of a multicultural Asian population

Mihir Gandhi[1,2,3*], Julian Thumboo[4,5], Nan Luo[6], Hwee-Lin Wee[7] and Yin-Bun Cheung[2,3]

Abstract

Background: There is conflicting evidence as to whether patients with chronic disease value hypothetical health states differently from individuals who have not experienced any long-lasting diseases. Furthermore, most studies regarding this issue have been conducted in western countries, with only one conducted in Asia. We aimed to evaluate possible systematic differences in the valuation of EuroQol Group five dimensions 3-level (EQ-5D-3L) health states by chronic disease patients and a population with no chronic disease in Singapore.

Methods: A face-to-face survey for the valuation of the 42 health states of the EQ-5D-3L using the visual analogue scale (VAS) method was conducted in Singapore. The survey also asked participants to report any chronic diseases they had. Ordinary least-square regression models were employed to assess possible differences in the valuation scores of all health states, severe health states and non-severe health states by individual chronic disease patient groups (diabetes, rheumatism, hypertension, heart diseases and lung diseases) and by a group of participants with no chronic disease. A difference of 4 to 8 points on the 100-point VAS was considered to be of practical significance.

Results: The analysis included 332 participants with at least one chronic disease and 651 participants with no chronic disease. After taking health state descriptors and covariates into account, mean valuation scores of the 42 health states by the heart disease group were higher by 4.6 points (p-value = 0.032) compared to the no chronic disease group. Specifically, the heart disease group valued severe health states 5.4 points higher (p-value = 0.025) than the no chronic disease group. There was no practically significant difference in the mean valuation score of non-severe health states between the heart disease group and the no chronic disease group. No practically significant differences were found in the mean valuation score of all health states, severe health states and non-severe health states between any other chronic disease group and the no chronic disease group.

Conclusions: In Singapore, heart disease patients valued EQ-5D-3L severe health states differently from individuals with no chronic disease. Other chronic disease groups did not value EQ-5D-3L health states differently from the no chronic disease group.

Keywords: Chronic disease, EQ-5D, Utility, Valuation

Background

There is conflicting evidence as to whether patients with chronic disease value their own health states differently from individuals who have not experienced any such diseases [1,2]. Similar conflicting results have been reported for how patients with chronic disease and individuals with no such disease experience value hypothetical health states [1,2]. The difference in valuation of health states between the patients and individuals with no disease experience may arise because the patients might have adapted to their condition or because individuals with no disease experience overestimate the impact of disease or disability on quality of life [3]. Most studies that have evaluated differences in valuations by these two groups have been conducted in western countries; only one study has reported on an Asian population [1,4]. This is important as there is

* Correspondence: mihir.gandhi@scri.edu.sg
[1]Biostatistics, Singapore Clinical Research Institute, Singapore, Singapore
[2]Centre for Quantitative Medicine, Duke-NUS Graduate Medical School, Singapore, Singapore
Full list of author information is available at the end of the article

evidence of meaningful differences between populations of different countries regarding the valuation of health states [5,6]. In addition, most of the studies that have compared the valuation by chronic disease patients and that of by a no chronic disease population have compared the valuation of only selected disease-related health states, without covering a range of mild to severe states. Only a few studies have investigated the potential systemic difference between valuation by specific chronic disease patients and individuals with no experience of chronic disease regarding health states with a wide range of severity [7,8].

Differences in valuation between chronic disease patients and individual with no chronic disease may affect the outcomes of analyses of healthcare interventions. Cost-utility analysis (CUA) is a cost-effectiveness analysis in which the effect of health-care interventions is measured in terms of quality-adjusted life-years (QALYs) gained. QALYs are estimated as the time spent in a health state (quantity of life) multiplied by its utility (quality of life). QALYs are also an important outcome for monitoring health status in individual patients, measuring population health and measuring the impact of health-care intervention in clinical studies [9]. The question of whose utility (general-population-derived or patient-derived) should be used in clinical decision making and economic evaluations of health-care interventions has been debated in the literature [10,11]. The answer depends on the purpose for which the utility is used and context in which it is used. A general population-derived utility is desirable when the utility is needed to inform decisions that allocate societal resources, while a patient-derived utility may be more appropriate when making treatment decisions guided by patient preferences. The Panel on Cost-Effectiveness in Health and Medicine in the United States and the National Institute for Health and Care Excellence in England and Wales recommend that a general-population-derived utility for health states be used for cost-effectiveness analyses [12,13]. However, the latest systematic review revealed that less than one-third of published CUAs use a general-population-derived utility; the remainder used a patient-derived utility, a clinician- or expert-derived utility, or authors' judgments [14]. Many investigators use a patient-derived utility because they believe that patients who have experienced the disease conditions can appraise their conditions more accurately than individuals who have not experienced such conditions [15]. On the other hand, CUAs using a general population-derived utility can help broader system-level decision making to prioritize health care funding in order to maximize the benefit for patients with different medical conditions- considering patients' as well as non-patients' perspectives [11]. This is a recommended approach when the health care is funded by the public/tax payers. However, if the health care costs are mostly paid by patients themselves, patient-derived utility should be considered. In Singapore, more than 60% of the health care costs are borne by patients [16]; and therefore patient-derived utility is relevant.

Utilities of health states from generic quality of life instruments, such as the EuroQoL Group five dimensions (EQ-5D) or Short Form six dimension (SF-6D), are preferred over health states from disease-specific quality of life instrument for CUAs. Utilities of generic health states allow comparisons of the effects on quality of life of different health-care interventions in different diseases. Currently, the EQ-5D is the most commonly used generic instrument for CUAs [14].

The present study draws on data from a valuation study of EQ-5D 3-level (EQ-5D-3L) health states in the Singapore general population which involves self-reporting of chronic diseases. We aimed to explore whether there are systematic differences in values for health states elicited by specific chronic disease patients (CDP) and by the no chronic disease population (NCDP). We also explored how the most severe health state and unconscious state were valued in relation to dead state by specific CDP and NCDP.

Methods

Valuation survey procedures

In 2009, the EQ-5D-3L—using the visual analogue scale (VAS) method—was used to conduct a cross-sectional, face-to-face survey of health state valuation in a representative sample of 1034 participants from the general population of Singapore. Singapore is a multi-ethnic city state with a rapidly increasing aging population. Its population is 75% Chinese, 13% Malay, 9% Indian (mostly Tamil speaking) and 3% others [17]. A multi-stage sampling approach was used to randomly select residential blocks, within which households were selected. We interviewed potential participants (one per household) who satisfied the pre-set recruitment quotas for ethnicity (400 Chinese, 400 Malay, and 234 Indians), gender (50% Female) and age (30% of 21–34 years, 40% of 35–49 years, and 30% of 50+ years). Within each ethnicity, there was a quota that half of the participants would use English for the interview and the remaining half would use their native language (i.e., Mandarin for Chinese, Malay for Malays and Tamil for Indians).

The EQ-5D-3L consists of 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) with 3 response levels for each dimension (1: no problems, 2: some problems, and 3: extreme problems). This instrument thus describes 243 health states. Each health state is represented by one response level from each of the 5 dimensions. For example, 11112 describes a health state with no problems on the first 4 dimensions and some problem related to anxiety/depression.

A subset of 42 EQ-5D-3L states was selected based on the protocol of Dolan [18]. Immediate death and the most severe state ('33333') were labeled as 'dead' and 'all-worst', respectively. Each participant was asked to compare between 'dying now' and 'living for the rest of his/her life in all-worst', from which the less desirable state was assigned a value 0 on the VAS. Each participant was then asked to value a unique set of 6 states from the subset of EQ-5D-3L states and either 'dead' or the 'all-worst' state, whichever one was not valued earlier at 0. The unique set of 6 health states that was assigned to each participant included states that were spread widely over the valuation space. A 100-point "feeling thermometer" with endpoints of 100 (most desirable, i.e., perfect health) and 0 (least desirable) was used as a VAS. For the six assigned states, participants were required to indicate where they would rate each of the states on the "feeling thermometer" by imagining themselves in that state for the rest of their life without changing. The participants were allowed to value more than one health state at the same level of VAS. In addition to the six assigned states, 'unconscious' state was also valued.

Participants were also asked to report their chronic diseases. The list of chronic diseases included diabetes, high blood pressure (hypertension), heart diseases, stroke, asthma or other lung diseases, cancer, rheumatism/back pain or other bone or muscle illness, mental illness (e.g., depression, anxiety neurosis, schizophrenia) and other illness (e.g., kidney problems or dialysis).

This study was approved by the SingHealth Centralized Institutional Review Board.

Analyses

The analysis included participants with diabetes, high blood pressure/hypertension, heart diseases, asthma/lung diseases, rheumatism/back pain/other bone-muscle illness, or no chronic disease. The number of participants with other chronic diseases was small (<10) and these participants were not included in the analyses described in this manuscript.

Participants who met the following criteria were excluded from our analysis: a) valued less than 3 health states, b) did not value 'dead' or 'all-worst' state, c) valued 'dead' or 'all-worst' or 'unconscious' state higher than all of the other states, d) gave the same valuation score to all the health states, e) self-reported or rated by the interviewers as having a poor understanding of the health states description or valuation tasks. The valuation score used in the analyses was 'raw' VAS valuation score, which ranged from 0 (worst possible score) to 100 (best possible score). There is no consensus among researchers or regulatory bodies regarding the optimal method of transforming the valuation scores into utility [19].

We performed a separate analysis to compare the valuations by participants in each of the CDP groups with those of the NCDP group. Each analysis included two ordinary least-square regression models. Model I was used for the comparison of overall difference in valuation scores (including all the health states) between each CDP group and the NCDP group. Model II was used for the comparison of the differences in valuation scores of non-severe health states and severe health states by including an interaction term between the indicator variable for severe health state (versus non-severe health state) and the indicator variable for the specific CDP group (versus the NCDP group). We considered health states with at least one dimension at level 3 as "severe" health states, and the remaining states as "non-severe".

Model I was performed for the valuation score with an indicator variable representing a specific CDP group, the members of which might have co-morbid conditions, versus the NCDP group as the independent variable. The model adjusted for indicator variables that represented the level of severity in each dimension of the health states. That is, including 2 indicator variables for each of the 5 dimensions of EQ-5D-3L. Furthermore, we included an indicator variable (commonly called 'N3' in the cost-utility analysis literature) to take into account additional disutilities when a severe problem (level 3) was reported on at least one dimension [20]. Finally, the comparison adjusted for ethnicity, gender, age, marital status, education level, religion and house type because the CDP group being analyzed and NCDP group might differ in these background characteristics.

Model II further included an interaction term between the CDP group indicator and the N3 in Model I. In this model, the coefficient of the CDP group provides an estimate of the difference between valuation scores of *non-severe health states*; whereas its sum with the interaction term provides an estimate of the difference between valuation scores of *severe health states* by the specific CDP group and the NCDP group after taking the health state descriptors and participants' background characteristics into account. Perfect health state, 'unconscious' state and 'dead' state were excluded from Models I and II. The perfect health state was assigned a valuation score of 100 for each participant. Since each of the participants valued 6 health states, we used the Eicker-Huber-White robust standard error for cluster data for statistical inference [21].

We compared the valuation of the 'all-worst' and 'unconscious' health states with the valuation of 'dead' state by each of the CDP groups and the NCDP group. The mean valuation scores of the 'all-worst' state and 'unconscious' state were compared with the 'dead' state using a paired t-test.

All the analyses were carried out using Stata/MP 10.1 for Windows. A minimally important difference of 4 to

8 points on the 100-point VAS was considered to be of practical significance [22-24].

Results

All 1034 participants provided demographic and health characteristic information. Nine participants had chronic diseases other than diabetes, high blood pressure/hypertension, heart disease, asthma/lung disease or rheumatism/back pain/other bone-muscle illness. Thirty-four participants valued 'dead' state higher than all the other states, 3 participants valued 'unconscious' state higher than all the other states, 1 participant did not value the 'all-worst' state and 4 participants were observed to have a poor understanding of valuation tasks. Hence, a total of 51 participants were excluded from the analysis. Table 1 shows the demographic and health characteristics of the 983 participants that were included in the analysis. The percentage of participants who had good-to-excellent self-reported general health varied between 68% and 75% among the CDP groups (Diabetes: 70/102, Rheumatism: 122/162, Hypertension: 107/145, Heart diseases: 29/44, Lung diseases: 30/44) as compared to 95% among NCDP group (621/651). Participants in the CDP groups were older, had lower education level and lower self-reported general health compared to participants in the NCDP group.

Table 2 summarizes the comparison of health state valuation scores between the CDP groups and the NCDP group. Mean observed differences between the CDP groups and the NCDP group regarding the valuation score of all the 42 EQ-5D health states, non-severe health states and severe health states ranged from −3.3 to 0.5, −3.7 to −1.0 and −2.8 to 2.1, respectively. After taking health state descriptors and covariates into account in the regression analysis, the mean differences between the CDP groups and the NCDP group regarding valuation scores of all the health states ranged from −2.5 to 1.6 (each p-value >0.05), except for the heart disease group. The adjusted mean valuation score of all the health states for the heart disease group was 4.6 points higher (95% CI: 0.4 to 8.9; p-value = 0.032) than that of the NCDP group. Similarly, after taking health state descriptors and covariates into account, the mean differences between the CDP group and the NCDP group regarding severe health state valuation scores ranged from −2.4 to 1.8 (each p-value >0.05), except for the heart disease group. The adjusted mean valuation score of severe states for the heart disease group was 5.4 points higher (95% CI: 0.7 to 10.1; p-value = 0.025) than that of the NCDP group. After taking health state descriptors and covariates into account, there was no practically significant difference in the mean valuation scores of non-severe health states between any CDP group and the NCDP group. The changes in the mean differences after the adjustment for the covariates could be due to differences in the distribution of demographic characteristics between the CDP and NCDP groups (please see Table 1). For example, the NCDP group had more participants married/living with partners, which was associated with higher value, compared with unmarried participants in multivariable analysis. After statistical adjustment, the difference between NCDP and lung disease groups would become smaller. Similarly, the NCDP group had differences in multiple demographic characteristics, such as more female participants, fewer Indian participants, and more participants following Buddhism/Taoism, which were associated with higher value, compared to the heart disease group. Thus, after statistical adjustment, the heart disease group had higher valuation score compared to the NCDP group. Other demographic characteristics did not have much influence on the valuation score (Details not shown).

Table 3 summarizes the comparison of valuation scores for the 'all-worst' state with those of the 'dead' state by disease group. Except for the heart disease group, the mean difference in valuation scores of 'all-worst' state and 'dead' state by the CDP groups and the NCDP group were within the range of −8.4 points (95% CI: −11.8 to −4.9; p-value < 0.001) to −5.3 points (95% CI: −8.8 to −1.8; p-value = 0.003). For the heart disease group, this difference was −2.3 points (95% CI: −7.3 to 2.8; p-value = 0.370).

Table 4 summarizes the comparison of valuation scores for 'unconscious' state with those of the 'dead' state by disease group. Except for the heart disease group, the mean difference in valuation scores of 'unconscious' state and 'dead' state by the CDP groups and the NCDP group were within the range of −0.1 points (95% CI: −3.0 to 2.7; p-value = 0.922) to 3.0 points (95%CI: −0.1 to 6.0; p-value = 0.057). For the heart disease group, this difference was 4.2 points (95% CI: 0.4 to 8.0; p-value = 0.030).

Discussion

We examined the potential effect of experience with chronic disease on the valuation of EQ-5D-3L health states using the VAS method in a multicultural Asian population. Valuation by participants with five different types of chronic disease (diabetes, rheumatism, hypertension, heart disease and lung disease) was compared with valuation by participants with no chronic disease.

The heart disease group valued the health states 5 points higher than did the NCDP group (p-value = 0.032), which is mainly attributed to the heart participants' valuation of the severe health states. This difference was statistically significant and larger than the minimal important difference of 4 points for the EQ-5D-3L valuation score, which indicates that the result is practically meaningful. The mean differences between the valuation by other CDP groups (diabetes, rheumatism, hypertension and lung diseases) and the NCDP group were all smaller than the minimal important difference.

Table 1 Demographic and health characteristics of the study participants

Characteristics	All participants (N = 983)	No chronic disease group (N = 651)	Diabetes (N = 102)		Rheumatism (N = 162)		Hypertension (N = 145)		Heart diseases (N = 44)		Lung diseases (N = 44)	
	n (%)	n (%)	n (%)	P-value*	n (%)	P-value*	n (%)	P-value*	n (%)	P-value*	n (%)	P-value*
Female	493 (50.2)	329 (50.5)	47 (46.1)	0.404	92 (56.8)	0.071	69 (47.6)	0.530	9 (20.5)	<0.001	29 (65.9)	0.044
Age (years)				<0.001		<0.001		<0.001		<0.001		0.048
21-29	190 (19.3)	168 (25.8)	0 (0.0)		7 (4.3)		2 (1.4)		0 (0.0)		14 (31.8)	
30-39	218 (22.2)	178 (27.3)	6 (5.9)		21 (13.0)		7 (4.8)		1 (2.3)		9 (20.5)	
40-49	261 (26.6)	181 (27.8)	22 (21.6)		43 (26.5)		24 (16.6)		7 (15.9)		5 (11.4)	
50-59	192 (19.5)	91 (14.0)	41 (40.2)		40 (24.7)		57 (39.3)		18 (40.9)		8 (18.2)	
60+	122 (12.4)	33 (5.1)	33 (32.4)		51 (31.5)		55 (37.9)		18 (40.9)		8 (18.2)	
Ethnicity				0.001		0.024		0.914		0.001		0.287
Chinese	363 (36.9)	234 (35.9)	26 (25.5)		67 (41.4)		56 (38.6)		9 (20.5)		16 (36.4)	
Malay	395 (40.2)	284 (43.6)	37 (36.3)		50 (30.9)		57 (39.3)		14 (31.8)		14 (31.8)	
Indian	225 (22.9)	133 (20.4)	39 (38.2)		45 (27.8)		32 (22.1)		21 (47.7)		14 (31.8)	
Education level				<0.001		<0.001		<0.001		<0.001		0.352
Primary (6 years) or less	187 (19.0)	74 (11.4)	48 (47.1)		60 (37.0)		58 (40.0)		22 (50.0)		12 (27.3)	
Secondary (11 years)	555 (56.5)	387 (59.5)	44 (43.1)		79 (48.8)		77 (53.1)		18 (40.9)		22 (50.0)	
Diploma/degree or higher	241 (24.5)	190 (29.2)	10 (9.8)		23 (14.2)		10 (6.9)		4 (9.1)		10 (22.7)	
Married/living with partner	739 (75.2)	481 (73.9)	84 (82.4)	0.090	128 (79.0)	0.233	120 (82.8)	0.022	38 (86.4)	0.106	23 (52.3)	0.001
Religion				0.051		0.094		0.796		0.017		0.894
Buddhism/Taoism	224 (22.8)	139 (21.4)	19 (18.6)		43 (26.5)		37 (25.5)		4 (9.1)		12 (27.3)	
Islam	410 (41.7)	289 (44.4)	41 (40.2)		53 (32.7)		63 (43.5)		17 (38.6)		16 (36.4)	
Hinduism/Sikhism	192 (19.5)	117 (18.0)	31 (30.4)		40 (24.7)		24 (16.6)		17 (38.6)		10 (22.7)	
Christianity	80 (8.1)	54 (8.3)	7 (6.9)		14 (8.6)		10 (6.9)		3 (6.8)		3 (6.8)	
No religion	77 (7.8)	52 (8.0)	4 (3.9)		12 (7.4)		11 (7.6)		3 (6.8)		3 (6.8)	
House type				0.377		0.175		0.786		0.204		0.533
Government owned: 4 rooms or smaller	668 (68.0)	449 (69.0)	71 (69.6)		106 (65.4)		96 (66.2)		25 (56.8)		27 (61.4)	
Government owned: 5 rooms or bigger	292 (29.7)	191 (29.3)	27 (26.5)		49 (30.3)		45 (31.0)		18 (40.9)		16 (36.4)	

Table 1 Demographic and health characteristics of the study participants *(Continued)*

Private	23 (2.3)	11 (1.7)	4 (3.9)	7 (4.3)	4 (2.8)	1 (2.3)	1 (2.3)	
General health status				<0.001	<0.001	<0.001	<0.001	<0.001
Excellent	97 (9.9)	85 (13.1)	3 (2.9)	5 (3.1)	2 (1.4)	2 (4.6)	2 (4.6)	
Very good	376 (38.3)	302 (46.4)	19 (18.6)	24 (14.8)	31 (21.4)	6 (13.6)	9 (20.5)	
Good	409 (41.6)	234 (35.9)	48 (47.1)	93 (57.4)	74 (51.0)	21 (47.7)	19 (43.2)	
Fair	93 (9.5)	30 (4.6)	27 (26.5)	36 (22.2)	32 (22.1)	13 (29.6)	11 (25.0)	
Poor	8 (0.8)	0 (0.0)	5 (4.9)	4 (2.5)	6 (4.1)	2 (4.6)	3 (6.8)	

*Comparison with the no chronic disease group using Fisher's exact test.

Table 2 Comparison of valuation of health states between the chronic disease groups and the no chronic disease group taking into account health state descriptors and covariates

Health States	Diabetes (n = 102)	Rheumatism (n = 162)	Hypertension (n = 145)	Heart diseases (n = 44)	Lung diseases (n = 44)	No chronic disease (n = 651)
All health states[1]						
Mean (SD)	43.6 (30.0)	43.4 (30.6)	44.4 (30.0)	43.1 (28.7)	40.6 (29.5)	43.9 (30.6)
Mean difference (95% CI)[2,3]	1.6 (−1.2, 4.3)	0.4 (−2.0, 2.8)	0.7 (−1.7, 3.1)	4.6 (0.4, 8.9)*	−2.5 (−6.2, 1.2)	-
Non-severe health states[1,4]						
Mean (SD)	71.6 (18.8)	71.3 (20.0)	71.2 (19.7)	69.8 (16.8)	69.0 (22.1)	72.7 (19.3)
Mean difference (95% CI)[2,3]	1.0 (−2.5, 4.6)	−0.3 (−3.4, 2.9)	−1.0 (−4.3, 2.3)	2.6 (−2.4, 7.6)	−2.6 (−9.0, 3.8)	-
Severe health states[1,4]						
Mean (SD)	31.4 (25.4)	31.9 (26.6)	33.6 (26.5)	33.7 (26.0)	28.7 (23.6)	31.5 (25.8)
Mean difference (95% CI)[2,3]	1.8 (−1.2, 4.7)	0.7 (−1.9, 3.3)	1.3 (−1.2, 3.9)	5.4 (0.7, 10.1)*	−2.4 (−6.1, 1.2)	-

[1]The study included 42 EQ-5D-3L health states, not including perfect health, unconscious and dead states. The perfect health state of EQ-5D-3L was assigned default value of 100 points on the visual analogue scale.
[2]Difference: mean scores of participants with chronic diseases minus mean scores of participants with no chronic disease.
[3]Using ordinary least square regression model adjusted for health state descriptors, disutility due to severe problems, ethnicity, gender, age, marital status, education level, religion and house type (see Methods section).
[4]EQ-5D-3L health states with at least one domain at severity level 3 are considered as 'severe' health states. Remaining health states are considered as 'non-severe' health states.
*P-value <0.05.

The 'all-worst' state (the most severe state of EQ-5D with all dimensions at extreme severity) was valued worse than the 'dead' state by the majority of participants across the different types of chronic diseases and the NCDP group. Except for the heart disease group, all CDP groups and the NCDP group valued the 'all worst' state statistically and practically significantly lower than the 'dead' state. The heart disease group's valuation of the 'all-worst' state was not statistically and practically significantly different from their valuation of the 'dead' state (difference = −2.3, p-value = 0.370).

We also found that the mean valuation score of the 'unconscious' state was likely to be equivalent to the 'dead' state by the NCDP group and all of the CDP groups except for the heart disease group. The difference in the mean valuation score of the 'unconscious' state and the 'dead' state by heart disease group was statistically significant and higher than the minimal important difference (difference = 4.2, p-value = 0.030), whereas the difference was statistically non-significant and less than 4 (minimal important difference) for the diabetes, rheumatism, hypertension, asthma/lung disease groups and the NCDP group.

A possible reason for heart disease patients giving higher valuation scores could be that a higher proportion of heart disease patients might have experienced one or more severe heath states, and this might have changed their perception regarding these health states. This might not be the case with other CDP groups and the NCDP group. On the other hand, the majority of CDP groups and the NCDP group might have experienced the non-severe health states, thus leading to their similar valuation of non-severe health states.

Wang et al. [5] in Singapore found that after adjusting for health state descriptors and demographic characteristics, there was no meaningful difference in the valuation of severe health states by diabetes patients and a population without diabetes. However, the study reported that diabetes patients valued the non-severe health states 13 points higher than did the no-diabetes population. Our

Table 3 Comparison of valuation scores between the 'all-worst' state and the 'dead' state by disease group

Valuation Scores	Diabetes (n = 102)	Rheumatism (n = 162)	Hypertension (n = 145)	Heart diseases (n = 44)	Lung diseases (n = 44)	No chronic disease (n = 651)
All-worst[#]						
Mean (SD)	3.3 (8.4)	2.1 (6.0)	3.1 (8.6)	4.8 (9.9)	1.7 (5.2)	3.1 (8.3)
Dead						
Mean (SD)	8.6 (13.3)	10.3 (15.9)	11.4 (17.1)	7.1 (10.4)	8.5 (10.9)	9.4 (14.1)
All-worst - Dead						
Mean difference (95% CI)	−5.3 (−8.8, −1.8)*	−8.1 (−11.0, −5.3)*	−8.4 (−11.8, −4.9)*	−2.3 (−7.3, 2.8)	−6.8 (−10.8, −2.7)*	−6.3 (−7.6, −4.9)*

[#]EQ-5L-3L health state with all its domains at severity level 3 is labelled as 'all-worst' health state.
*P-value < 0.05.

Table 4 Comparison of valuation scores between the 'unconscious' state and the 'dead' state by disease group

Valuation Scores	Diabetes (n = 102)	Rheumatism (n = 167)	Hypertension (n = 147)	Heart diseases (n = 46)	Lung diseases (n = 44)	No chronic disease (n = 667)
Unconscious						
Mean (SD)	11.6 (11.1)	10.1 (12.3)	12.0 (13.0)	11.3 (11.7)	10.6 (9.6)	11.8 (12.6)
Dead						
Mean (SD)	8.6 (13.3)	10.3 (15.9)	11.4 (17.1)	7.1 (10.4)	8.5 (10.9)	9.4 (14.1)
Unconscious - Dead						
Mean difference (95% CI)	3.0 (−0.1, 6.0)	−0.1 (−3.0, 2.7)	0.6 (−2.6, 3.8)	4.2 (0.4, 8.0)*	2.1 (−1.4, 5.7)	2.4 (1.2, 3.6)*

*P-value < 0.05.

findings do not fully support their results. It should be noted that Wang et al. included only 3 non-severe health states; hence their findings have limited applicability. On the other hand, we used 14 non-severe health states, which represent more generalized findings.

Our study findings are consistent with those of Pickard et al. [8]. Using the time trade-off method, Pickard et al. found no meaningful difference in valuation scores between CDP (arthritis, diabetes, depression, hay fever, cancer) and NCDP, except for heart failure patients [8]. Pickard et al. found that after adjusting for covariates, patients with heart failure only, and patients with heart failure and at least one other chronic disease, gave valuation scores higher by 25 points (n = 6, p-value = 0.222) and 7 points (n = 129, p-value = 0.049), respectively, compared to NCDP.

A possible explanation for no practical differences in the mean valuation score between individuals with chronic diseases and individuals without any chronic diseases might be because in this exercise, individuals with and without chronic diseases are valuing many hypothetical health states that are unlikely to reflect the actual health state(s) that one has experienced. As such, it is probably not surprising that generally speaking, individuals with chronic disease might value them similarly to individuals with no chronic disease.

This study has several potential limitations. First, the chronic disease conditions were self-reported by the participants. We did not collect any further information to confirm the disease, the severity of the disease or the time spent with the disease. Hence, there could be a chance of misclassification regarding reported diseases. A Finnish study showed that the sensitivity and specificity of self-reported chronic diseases (diabetes, hypertension, coronary heart disease, asthma and rheumatoid arthritis) could range from 78% to 96% and 96% to 99%, respectively [25]. This indicates a relatively large possibility that patients with chronic diseases could be misclassified into the NCDP group, but a small possibility that those with no chronic disease could be misclassified into a chronic disease group. This should mean that the difference between CDP and NCDP might be under-estimated but not over-estimated. Furthermore, this is a secondary analysis of existing data. The limited information related to disease conditions does not allow us to investigate any concrete reasons for the differences or lack of differences between the CDP and NCDP. Second, although our study had a sizable CDP group, nearly 80% of the CDP group self-reported their health status as good to excellent. Hence, our study findings may not be generalized to patients at a severe or unstable stage of chronic disease. Third, our study included only five chronic diseases (with a relatively small number of participants) and only one life-threatening chronic disease (heart diseases). Thus, our study findings may not be assumed to generalize to other life-threatening chronic diseases. Fourth, we performed separate statistical tests for comparing valuation by each CDP group with valuation by the NCDP group without multiplicity adjustment. Furthermore, the sample size was not powered for this analysis. Thus, the statistically significant findings might be due to inflated Type I error and therefore require further confirmation. Nevertheless, our findings are based on a random sample of a chronic disease population from the Asian general population. It also included many health states with a wide range of severity. It also has potential to generalize the findings for non-life-threatening chronic disease patients. We encourage conducting a larger study that includes a greater variety of life-threatening chronic disease patients, as well as varying severity levels and the verification of disease conditions and severity.

Conclusions

Our study findings suggest that heart disease patients value severe EQ-5D-3L health states differently than individuals who have no experience with chronic disease when analyzed using the VAS method in a Singaporean population. However, the experience of chronic diseases other than heart disease does not necessarily result in a higher or lower valuation across all the health states of EQ-5D-3L.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
JT conceived the study, and participated in the study design and coordination. HLW and LN participated in the study design and coordination. YB participated in the design of the study, statistical analysis and interpretation of the results. MG carried out the data analysis and interpretation, and wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements
The study was funded by a programme grant 03/1/27/18/226 from the Biomedical Research Council of Singapore. The last author (YBC) was supported by Singapore Ministry of Health's National Medical Research Council under its Clinician Scientist Award. We acknowledge the medical editing assistance of Jon Kilner, MS, MA (Pittsburgh, Pennsylvania, USA).

Author details
[1]Biostatistics, Singapore Clinical Research Institute, Singapore, Singapore. [2]Centre for Quantitative Medicine, Duke-NUS Graduate Medical School, Singapore, Singapore. [3]Department of International Health, School of Medicine, University of Tampere, Tampere, Finland. [4]Department of Rheumatology and Immunology, Singapore General Hospital, Singapore, Singapore. [5]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [6]School of Public Health, National University of Singapore, Singapore, Singapore. [7]Department of Pharmacy, Faculty of Science, National University of Singapore, Singapore, Singapore.

Received: 9 August 2014 Accepted: 23 December 2014
Published online: 23 January 2015

References

1. Peeters Y, Stiggelbout AM. Health state valuations of patients and the general public analytically compared: a meta-analytical comparison of patient and population health state utilities. Value Health. 2010;13(2):306–9.
2. Dolders MG, Zeegers MP, Groot W, Ament A. A meta-analysis demonstrates no significant differences between patient and population preferences. J Clin Epidemiol. 2006;59(7):653–64.
3. Bremner KE, Chong CA, Tomlinson G, Alibhai SM, Krahn MD. A review and meta-analysis of prostate cancer utilities. Med Decis Making. 2007;27:288–98.
4. Wang P, Tai ES, Thumboo J, Vrijhoef HJ, Luo N. Does Diabetes Have an Impact on Health-State Utility? A Study of Asians in Singapore. Patient 2014 [Epub ahead of print]
5. Luo N, Wang P, Thumboo J, Lim YW, Vrijhoef HJ. Valuation of EQ-5D-3L health states in Singapore: modeling of time trade-off values for 80 empirically observed health states. Pharmacoeconomics. 2014;32(5):495–507.
6. Johnson JA, Luo N, Shaw JW, Kind P, Coons SJ. Valuations of EQ-5D health states: are the United States and United Kingdom different? Med Care. 2005;43:221–8.
7. Krabbe PF, Tromp N, Ruers TJ, van Riel PL. Are patients' judgments of health status really different from the general population? Health Qual Life Outcomes. 2011;9:31.
8. Pickard AS, Tawk R, Shaw JW. The effect of chronic conditions on stated preferences for health. Eur J Health Econ. 2013;14(4):697–702.
9. Kind P, Lafata JE, Matuszewski K, Raisch D. The use of QALYs in clinical and patient decision-making: issues and prospects. Value Health. 2009;12 Suppl 1:S27–30.
10. McTaggart-Cowan H. Elicitation of informed general population health state utility values: a review of the literature. Value Health. 2011;14(8):1153–7.
11. Stamuli E. Health outcomes in economic evaluation: who should value health? Br Med Bull. 2011;97:197–210.
12. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the panel on cost-effectiveness in health and medicine. JAMA. 1996;276:1253–8.
13. National Institute for Health and Clinical Excellence (NICE): Guide to the Methods of Technology Appraisal. NICE; 2013 [http://www.nice.org.uk/article/pmg9/resources/non-guidance-guide-to-the-methods-of-technology-appraisal-2013-pdf]
14. Brauer CA, Rosen AB, Greenberg D, Neumann PJ. Trends in the measurement of health utilities in published cost-utility analyses. Value Health. 2006;9(4):213–8.
15. Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. JAMA. 1996;276(14):1172–7.
16. World Health Organization. World Health Statistics 2014. WHO Press. [Internet] http://apps.who.int/iris/bitstream/10665/112738/1/9789240692671_eng.pdf. Accessed November 14, 2014
17. Singapore Department of Statistics. Census of Population 2010. Statistical Release 1. Demographic characteristics, education, language and religion. Singapore: Department of Statistics 2011 [Internet]. http://www.singstat.gov.sg/docs/default-source/default-document-library/publications/publications_and_papers/cop2010/census_2010_release1/cop2010sr1.pdf. Accessed April 16, 2014
18. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997;35(11):1095–108.
19. Lamers LM. The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value sets. Med Care. 2007;45(3):238–44.
20. Szende A, Oppe M, Devlin N. EQ-5D value sets: inventory, comparative review and user guide (EuroQol group monographs, Vol. 2). The Netherlands: Springer; 2007.
21. Williams R. A note on robust variance estimation for cluster-correlated data. Biometrics. 2000;56:645–6.
22. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14:1523–32.
23. Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health Qual Life Out. 2007;5:70.
24. Luo N, Johnson JA, Coons SJ. Using instrument-defined health state transitions to estimate minimally important differences for four preferencebased health-related quality of life instruments. Med Care. 2010;48:365–71.
25. Oksanen T, Kivimäki M, Pentti J, Virtanen M, Klaukka T, Vahtera J. Self-report as an indicator of incident disease. Ann Epidemiol. 2010;20(7):547–54.

PUBLICATION III

Comparison of health state values derived from patients and individuals from the general population

Mihir Gandhi, Ru San Tan, Raymond Ng, Su Pin Choo, Whay Kuang Chia, Chee Keong Toh, Carolyn Lam, Phong Teck Lee, Nang Khaing Zar Latt, Kim Rand-Hendriksen, Yin Bun Cheung, Nan Luo

Quality of Life Research 2017, 26(12):3353-3363
https://doi.org/10.1007/s11136-017-1683-5

Comparison of Health State Values Derived from Patients and Individuals from the General Population

Mihir Gandhi[1,2,3], Ru San Tan[4], Raymond Ng[5], Su Pin Choo[5], Whay Kuang Chia[5], Chee Keong Toh[5], Carolyn Lam[4], Phong Teck Lee[4], Nang Khaing Zar Latt[4], Kim Rand-Hendriksen[6,7], Yin Bun Cheung[1,2,3], Nan Luo[8]*

[1]Department of Biostatistics, Singapore Clinical Research Institute, #02-01, 31 Biopolis Way, Singapore

[2]Centre for Quantitative Medicine, Duke-NUS Medical School, Level 6, Academia, 20 College Road, Singapore

[3]Tampere Center for Child Health Research, University of Tampere and Tampere University Hospital, Arvo Building, Lääkärinkatu 1, Tampere, Finland

[4]Department of Cardiology, National Heart Centre Singapore, 5 Hospital Drive, Singapore

[5]Division of Medical Oncology, National Cancer Centre Singapore, 11 Hospital Drive, Singapore

[6]Health Services Research Centre, Akershus University Hospital, Forskningsveien 3A, 0373 Oslo, Norway

[7]Deptartment of Health Management and Health Economics, University of Oslo, Forskningsveien 3A, 0373 Oslo, Norway

[8]Saw Swee Hock School of Public Health, National University of Singapore, Tahir Foundation Building (MD1), 12 Science Drive 2, Singapore

Email addresses of co-authors:

MG: mihir.gandhi@scri.edu.sg (ORCID: 0000-0002-8902-2710)

RST: tan.ru.san@singhealth.com.sg (ORCID: 0000-0003-2086-6517)

RN: raymond.ng.c.h@singhealth.com.sg (ORCID: 0000-0001-9651-234X)

SPC: choo.su.pin@singhealth.com.sg (ORCID: 0000-0002-8925-3922)

WKC: john.chia.w.k@singhealth.com.sg (ORCID: 0000-0003-4198-9831)

CKT: toh.chee.keong@singhealth.com.sg (ORCID: 0000-0002-5691-6144)

CL: carolyn.lam.s.p@nhcs.com.sg (ORCID: 0000-0003-1903-0018)

PTL: phongteck.lee@mohh.com.sg

NKZL: nangkhaingzar.latt@mohh.com.sg

KR: kim.rand-hendriksen@medisin.uio.no (ORCID: 0000-0001-7692-4099)

YBC: yinbun.cheung@duke-nus.edu.sg (ORCID: 0000-0003-0517-7625)

* Corresponding author:

Nan Luo

Associate Professor, Saw Swee Hock School of Public Health

National University of Singapore, 12 Science Drive 2, Singapore

Email: ephln@nus.edu.sg

Telephone: +65-65164966

ORCID: 0000-0001-7980-6979

Running head: Health state values of patients and public

Funding: The study was supported in part by the EuroQol Research Foundation (EQ Project 2015220), The Netherlands, and in part by the Duke-NUS Signature Research Program (WBS R-913-200-040) funded by the Agency for Science, Technology and Research (A*STAR), Singapore, and the Ministry of Health, Singapore. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

ABSTRACT

Purpose: Utility values are critical for cost-utility analyses that guide healthcare decisions. We aimed to compare the utility values of the 5-level EuroQoL-5Dimension (EQ-5D-5L) health states elicited from members of the general public and patients with heart disease or cancer.

Methods: In face-to-face interviews with 157 heart disease patients, 169 cancer patients, and 169 members from the general population, participants valued 10 EQ-5D-5L health states using a composite Time Trade-Off method.

Results: Pooling utility values for all health states, heart disease patients and cancer patients had mean utility values lower by 0.11 points (P-value=0.014) and 0.06 points (P-value=0.148), respectively, compared to the general population. Adjusting for sociodemographic characteristics, differences in health state utility values between the patient and the general populations were rendered non-significant, except that heart disease patients gave higher utility values (mean difference=0.08; P-value=0.007) to mild health states than the general population. Difference in utility values, defined as utility value of a better health state minus that of a poorer health state, was higher among heart disease patients compared to the general population, before and after adjusting for sociodemographic characteristics.

Conclusions: Patients may differ from members of the general population in the strength of their preferences for hypothetical health states. Using utility values derived from the general population may under-estimate the comparative effectiveness of healthcare interventions for certain diseases, such as heart diseases.

Keywords: EQ-5D; time trade-off; cancer; heart disease; utility; preference

Introduction

Health state utility values are usually elicited from either the general population or patients [1-3]. General population-derived utility values have been recommended for use in cost-effectiveness analysis to inform decisions involving allocation of societal resources [4]. Health utility instruments such as the EuroQol 5 dimensions (EQ-5D) are designed to generate health state utility values based on the health preferences of the general population. Members of the general population, however, may have different health preferences compared to patients, and they may lack understanding of what it means to live in impaired health [5]. Therefore, patient-derived utility values are desirable for effectiveness analysis to inform clinical decision making from a patient-centered perspective.

The issue of whose preferences to use is only important insofar as health preferences differ between the populations. Some studies indicate that there are important differences between patient-derived and general population-derived health state values [6-10]. For example, a meta-analysis found that utility values elicited from patients tend to be higher than those elicited from the general population [8]. However, other studies, including a meta-analysis, report no or minimal difference in the utility values elicited from general and patient populations [9, 11-13]. The mixed findings could be due to multiple factors, including type of condition, severity of condition, and valuation method. For example, one study reports that health state values derived using the Visual Analog Scale (VAS) method were similar between individuals with and without arthritis, but were different between individuals with and without heart disease [9]; utility values elicited from patients with arthritis and the general population using the VAS method were similar but values elicited from the two populations using the Time Trade-Off (TTO) method were different [14]. The mixed findings may also be due to the fact that some of the studies were underpowered, poorly designed, or poorly executed [8].

In this study, we investigated the impact of chronic diseases on measurement of the utility of health states defined by the 5-level EuroQoL-5Dimension (EQ-5D-5L) questionnaire [15]. The EQ-5D-5L is a new version of EQ-5D, and has demonstrated better measurement properties than the 3-level EQ-5D (EQ-5D-3L) [16, 17]. The primary objective of this study was to compare two patient populations, heart disease and cancer patients, with the general population. The two diseases were selected because of their high disease burden globally [18, 19]. In all three groups, we elicited utility values for a set of 10 EQ-5D-5L health states using the TTO method. Both TTO and EQ-

5D are recommended by the National Institute for Health and Care Excellence (NICE) for generating utility values in economic evaluations of health technologies [20].

Methods

Participants

Heart disease and cancer patients attending outpatient clinics at the National Heart Centre Singapore (NHCS) and the National Cancer Centre Singapore (NCCS), respectively, were invited to participate during their routine visits. The NHCS and NCCS are the largest capacity specialty centers in Singapore for cardiovascular disease and cancer patients respectively. The main eligibility criterion for the heart disease patients was hospitalization primarily for a heart disease such as coronary heart disease or heart failure treatment in the last five years; this criterion was used to screen out patients with mild heart conditions. The main eligibility criterion for the cancer patients was to have histologically confirmed cancer of any type and stage in the last five years; this criterion was used to screen out cancer survivors. The study also included a sample from the general population recruited from three shopping malls in Singapore. All participants were between 21 and 80 years old, able to read and communicate in English or Chinese (Mandarin), and well enough for an interview. A quota sampling based on age and gender distributions similar to the Singapore census was used to generate the general population sample.

Valuation interview

The study design was cross-sectional, and consenting participants were interviewed face-to-face in a computer-assisted interviewing protocol. All interviews were conducted by the same trained interviewer using the EuroQol Valuation Technology (EQ-VT) computer program running from a laptop in either English or Chinese according to participant's preference [21]. The interviews with the patients were conducted in the hospitals in a quiet waiting area; the general population participants were interviewed in a quiet place of the malls where they were recruited. The study was approved by the SingHealth Centralized Institutional Review Board.

The interviewer followed a standard interviewer script in all interviews [22]. Both the Chinese and English versions of the EQ-VT computer program and the interviewer script were tested among Singaporean general population as well as patient populations in pilot studies. The valuation tasks using the EQ-VT computer program were well

understood and accepted by all three major local ethnicities (Chinese, Malay and Indian) in the general and patient populations [21]. The advantages of using a computer program include reduced interviewer burden, inter-interview variation, and errors and violation of interview protocol. The EQ-VT program is designed to collect data on the processes of each interview and upload the data daily to the server. By analyzing the process data, we were able to identify any errors the interviewer made and intervene timely whenever necessary.

The interviews started with some warm-up questions asking the participants to describe their own health using the EQ-5D-5L questionnaire. Subsequently, the TTO-based valuation task was explained to participants using the state of "in a wheelchair" as an example, after which three practice EQ-5D-5L health states were administered to familiarize participants with the task and EQ-5D-5L health states of varying severity. The practice states were followed by TTO valuation of a 10 EQ-5D-5L health states. The interviews ended with some feedback and background questions.

A detailed description of the TTO and the EQ-VT protocol can be found elsewhere [23]. Briefly, the objective of the task was to identify the point of preferential indifference between 10 years of life in the described target state, followed by death, and a shorter life ($x \leq 10$ years) in full health, followed by death. With a defined utility value of 1 for 10 years in full health, the utility value of the target state can be calculated as x/10. For states considered to be worse than death (respondent preferred a life of 0 to 10 years in the target state), a lead-time of 10 years was added to both alternatives in order to elicit a negative utility value for the state. The utility value of a worse than death health state was calculated as (x-10)/10 such that the utility value of each health state is bounded at -1 and 1; 0 represents value for the 'dead' state.

In addition to the valuation interview, clinical information was collected from patients by interview or directly from their medical records. Clinical information included diagnoses, year of diagnosis, and clinical assessment such as New York Heart Association (NYHA) functional classification and Canadian Cardiovascular Society (CCS) functional classification of angina for heart disease patients; and cancer stage and Eastern Cooperative Oncology Group (ECOG) performance status for cancer patients. All participants were also asked to self-report their current and past chronic diseases.

The EQ-5D-5L health states

In this study, all participants were asked to value the same set of 10 health states, in random order. All health states were defined using the EQ-5D-5L system which contains five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) and five functional levels for each dimension (broadly corresponding to no problems, slight problems, moderate problems, severe problems, and extreme problems). EQ-5D-5L health states are conventionally described using a 5-digit index, where the digits represent the functional level of each dimension in the conventional order of presentation (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). For example, the health state '11122' indicates slight problems (level 2 severity) in both pain/discomfort and anxiety/depression and no problems in mobility, self-care, and usual activities. The 10 health states were 11122, 21121, 21222, 21232, 32232, 32333, 22224, 31242, 53343, and 33453. The first three health states (11122, 21121, 21222) with all dimensions at severity level either 1 or 2 were considered as 'mild' health states, the last four health states (22224, 31242, 53343, 33453) with at least one dimension at severity level either 4 or 5 were considered as 'severe' health states, and the remaining three health states (21232, 32232, 32333) were considered as 'moderate' health states.

Statistical analyses

The planned sample size for the study was 525 participants (175 heart disease patients, 175 cancer patients and 175 participants from the general population). The sample size was estimated to detect a difference of 0.1 standardized effect size in mean value of the 10 elicited health states from a group of patients and general population using a two-sided test for 5% type-I error rate and 80% power, assuming 10% participants may provide logically inconsistent valuation or do not complete the interviews.

Participants who met the following criteria were excluded from the analysis: a) gave the same utility value to all the health states, b) gave negative or zero utility value to all health states (i.e., considered all health states worse than or equal to death). The use of stricter logical inconsistency criteria (i.e., further excluding participants who valued one or more mild states as equal to or lower than any severe states) was explored. It led to exclusion of more participants but had no material impact on results and therefore was not adopted.

We performed ordinary least-square (OLS) regression analysis to compare the mean utility values elicited from the patient groups with those of the general population group. The regression model was performed separately for each of the 10 health states, across 10 health states, the three mild health states (11122, 21121, 21222), the three moderate health states (21232, 32232, 32333), and the four severe health states (22224, 31242, 53343, 33453). Finally, these models were repeated by adjusting for ethnicity, gender, age, marital status, education level, employment status, religion and household income level because the patient groups and the general population group differed in demographic and socioeconomic characteristics.

Lastly, we compared the differences in utility values between the health states as elicited from members of the general population, heart disease patients and cancer patients. As the difference in utility values of two health states is usually used to approximate utility gained from transitions between the two health states, this comparison helped us to assess whether the different sets of utility values would give similar estimates when used to determine utility gained from health state transitions. First, differences in utility were calculated separately for the heart disease patient group, cancer patient group, and the general population group for 35 pairs of health states where one is no worse than the other in any of the five domains. Second, differences in the differences between a patient group and the general population group were presented using a line graph and tested by OLS regression models. The graph also presented the differences in differences adjusted for participants' demographic and socioeconomic characteristics, calculated using the regression models performed separately for each pair of health states.

All the models involving more than one utility value per participant used the Eicker-Huber-White robust standard error for cluster data for statistical inference [25]. We also performed both unadjusted and adjusted analysis using mixed-effect models with participant specific-intercepts. The results were very similar to those based on OLS models with robust standard error. Hence, results of only OLS models were presented. All the analyses were carried out using Stata/MP 13.1 for Windows. A minimally important difference of 0.05 points for the utility values was considered to be of practical significance [26].

Results

Of 525 participants who completed the interview, 30 were excluded from analyses: 24 participants assigned the same value to all 10 health states, and six participants valued all the health states worse than or equal to death. There

was no systematic difference in demographic and health characteristics between participants included and excluded from the analysis, except that slightly more heart disease patients (n = 18) were excluded compared to the general population participants (n = 6) and cancer patients (n = 6) (Appendix Table 1). Table 1 shows the demographic, socioeconomic and health characteristics of the 495 participants included in the analyses (169 in the general population group, 157 in the heart disease patients group, and 169 in the cancer patients group). Participants in the patient groups were older, had lower education level, and poor self-reported general health compared to participants in the general population group. A total of 62 (37%) participants in the general population group self-reported to have one or more chronic diseases, such as hypertension (15%), hyper/dyslipidemia (12%), and lung diseases (10%). The majority of demographic characteristics of the general population sample were similar to the census population (Appendix Table 2) [27].

The majority of the heart disease patients reported no breathlessness with heavy or moderate exertion (95%), and had coronary artery disease (64%), atrial fibrillation (34%), arrhythmias (29%), myocardial infraction (28%), or heart failure (20%). Mean time from the most recent episode of symptom onset to the interview date was 1.2 years (SD 1.5). The majority of the heart disease patients (85%) had other chronic diseases such as hypertension (57%), hyper/dyslipidemia (59%), and diabetes mellitus (29%). Mean time from the most recent hospitalization due to a heart disease problem to the interview date was 1.4 years (SD 2.0). (Appendix Table 3)

The majority of the cancer patients reported ECOG performance status 0-1 (99.4%), and had breast cancer (29%), colorectal cancer (21%), or lymphoma (11%). The percentage of patients in cancer stages 0-1, 2-3, and 4 were 21%, 42% and 36% respectively. Mean time from the diagnosis of the most recent cancer to the interview date was 1.9 years (SD 1.4). (Appendix Table 4)

Only 5% of patients in the heart disease group self-reported to have had cancer. Similarly, 4% of patients in the cancer group have had heart disease, and less than 5% of participants in the general population group had either heart disease or cancer.

Table 2 summarizes the comparison of utility values between the patient groups and the general population group. Pooling all 10 states, heart disease patients displayed a mean utility value that was lower by 0.11 points (P-value=0.014) than the general population group. Heart disease patients valued moderate and severe health states

lower by 0.05 points (P-value=0.353) and 0.25 points (P-value <0.001), respectively, than members of the general public. There was no statistically significant difference in mean utility value for mild health states between the two groups (difference=0.03). After adjusting for demographic and socioeconomic variables in the regression models, the overall difference weakened and became statistically non-significant. However, the mean utility value based on heart disease patients was lower by 0.05 (P-value=0.534) for severe health states, and higher by 0.08 points (P-value=0.007) and 0.09 points (P-value=0.176) for mild and moderate health states, respectively, than that based on the general population members.

Unlike heart disease patients, there was no statistically significant difference in mean utility value for all health states including mild, moderate and severe health states between cancer patient group and the general population group. After taking the effects of demographic and socioeconomic characteristics into account, the mean differences in utility values between the two groups were smaller than the ones without adjustment and statistically non-significant (Table 2). Among all the covariates adjusted for the analysis, age had strongest association with utility (Wald test P <0.0001; see Appendix Table 4). Older age was associated with lower utility value.

Mean (unadjusted and adjusted) differences in differences between selected health states for the heart disease patients versus the general population group are presented in Fig. 1. It shows that the covariate-adjusted differences in differences were higher than 0.05, a possibly minimally important difference for utility values, for 21 of 35 pairs of health states (60%; P-value <0.05 for 11 pairs), and lower than 0.05 for 8 pairs of health states (23%; P-value ≥0.05 for all) in heart disease patients than the general population group. Similar analysis comparing cancer patients and the general population group (Fig. 2) showed that adjusted differences in differences were greater than 0.05 for 10 of 35 pairs of health states (29%; P-value ≥0.05 for all), and lower than 0.05 for 3 pairs (9%; P-value ≥0.05 for all).

Discussion

Overall, both heart disease and cancer patients tended to value the utility of EQ-5D-5L defined health states lower than members of the general public. The difference between cancer patients and the general public was consistent across health states of different severity, and disappeared after adjusting for the differences in demographics. In contrast, the difference between heart disease patients and the general public depended on health state severity, and

demographics explained the difference in severe states but did not fully account for the difference in mild and moderate states. These results suggest that whether the health preferences of patients and members of the general population differ depends on many factors.

So far two studies, conducted by Pickard et al. [13] and Gandhi et al. [9], compared valuation of 3-level EQ-5D (EQ-5D-3L) health states between heart disease patients and the general population. Our study is consistent with the two studies in that heart disease patients give mild and moderate EQ-5D-3L health states higher values compared to the general population, which could be explained by the theory of adaptation or coping. Patients are more likely to have experienced and adapted to mild and moderate health problems defined by EQ-5D-3L than healthy individuals. As a result, patients do not perceive those health states as intolerable or undesirable as the general public. A recent study of patients with type 2 diabetes mellitus reported similar results [6], suggesting that higher valuation of mild health problems may be a common phenomenon in some patient populations. On the other hand, our finding that heart disease patients give lower values to severe health states than the general population was not observed in previous studies. For example, Gandhi et al. [9] found the opposite – heart disease patients valued severe health states higher than the general public, which can be explained by adaptation. However, the adaptation theory could also explain low valuation of severe health problems. If the severe health problems are difficult to adapt to and their detrimental effects are also difficult to imagine, patients who had the experience could perceive them more undesirably than individuals who never had such experience [28]. Coincidentally, a recent study found that, compared to the general population, breast cancer and rheumatoid arthritis patients valued EQ-5D-5L defined mobility and self-care problems less undesirable but pain/discomfort and anxiety/depression more undesirable [29]. In another study comparing value of depression states between individuals with and without depression has shown that individual with depression valued depression lower than the individual who have not experienced it [30]. These studies might be a good support to the theory we used to explain the low valuation of health problems by patients, considering these two facts: 1), physical problems such as partial paralysis are easier to adapt than sensational problems such as pain and depression if they are persistent; 2), three of the four severe health states we used in our study involved severe or extreme pain/discomfort or anxiety/depression and only one health state involved extreme mobility problems (i.e., unable to walk about). The reasons for not observing this result in previous studies could be due to use of small sample size as well as health states characterized mainly by physical health problems.

Pickard et al. [13] also compared valuation by cancer patients and the general population. It showed results similar to our study that cancer patients give slightly lower but statistically insignificant values than the general population. A study conducted by Krabbe et al. [14] showed that cancer patients give a higher value to EQ-5D-3L health states than the general population. However, the comparison did not adjust for the effects of demographic and socioeconomic characteristics; the patients were significantly older than the general population group in that study. It can be reasonably postulated that utility values derived from cancer patients and the general population could be similar if the effect of age had been adjusted for in that study. These consistent findings suggest that cancer patients and the general population may have very similar preferences for EQ-5D-5L health states. Therefore, the general population-based EQ-5D-5L value sets could be sufficient for measuring the utility of various health outcomes when the measurement should be based on the preferences of cancer patients such as in clinical decision making.

There are a few possible reasons for why heart disease and cancer patients were different in valuation of EQ-5D-5L health states compared to the general population. For severe health states, it could be due to different levels of adaptation. It is possible that cancer patients adapt better to such health states than heart disease patients given that most of them have received surgery and/or chemotherapy, both of which have huge impact on patients' health and quality of life. For mild and moderate health states, cancer patients might focus more on their impact on quality of life than heart disease patients and therefore they are willing to trade a slightly larger portion of their life expectancy (10 years in this study). It is also possible that, compared to members of the general public and heart disease patients of the same age, cancer patients are more likely to perceive their actual life expectancy to be shorter than 10 years and therefore they are more generous in trading the 'extra' life years in the given life expectancy. This trading behavior would lead to lower utility values for health states of all severities. It could offset the effect of adaptation and as a result make cancer patients similar to the general population in valuation of health outcomes using the TTO method. Our finding that older age is associated with lower utility value in this study could be evidence for the effect of self-perceived life expectancy on TTO-based valuation using a standardized timeframe. Further studies are needed to test the generalizability of the findings on the population/disease-specific valuation outcomes and ascertain the underlying valuation behaviors.

Our analysis of the differences in differences suggests that utility values derived from the general population will lead to smaller differences than utility values derived from heart disease patients, if those are used to determine

utility gained from transitions between the studied health states. This means that the general population-based EQ-5D-5L value sets are less likely underestimate treatment benefits for heart disease patients if they are used in clinical trials or other longitudinal studies. Therefore, it may be worthwhile to develop a heart disease patients based EQ-5D-5L value set for use in studies for informing clinical decision making or other decision making in which patients' health preferences are most relevant. One the other hand, differences in utility between EQ-5D-5L health states based on utility values derived from cancer patients and the general public were found to be quite similar, suggesting that the general population based EQ-5D-5L value sets may be sufficient for use in clinical decision making for cancer patients. The differing results for heart diseases and cancers in this analysis suggest that patients-based EQ-5D-5L value sets, if needed for evaluating treatments, should be disease specific.

Our finding that older age is associated with lower health-state valuation is in concordance with findings of previous studies using a similar valuation method [31, 32]. The effect of age on TTO values could be due to the different life experience and perceived responsibilities of young and old individuals. For young individuals, they may tend to look forward to more life experience and responsibilities in the future and therefore prefer longer life to better quality of life; for old individuals, they are likely to have experienced a lot and fulfilled their responsibilities in past lives and are willing to trade life years for a healthy life. Further studies are needed to test the generalizability of the finding in other patient populations.

Our study has several potential limitations. First, our study sample did not include inpatients. As inpatients are likely to be in worse functional status which was associated with lower utility values in heart disease patients in our study (data not shown), the difference in utility values between the heart disease patients and the general population might be under-estimated. Nevertheless, it would be difficult to conduct the very cognitively demanding valuation tasks with inpatients. Second, our study involved only cardiovascular diseases and cancers. Therefore, our specific study findings may not be generalizable to other patient populations. However, the observation that clinical groups differ in terms of how they value health is likely to be generalizable. Third, our general population sample was recruited from shopping malls. As shopping is generally a pleasant leisure activity which could have an impact on valuation of health, the magnitude of the difference observed between the patient populations and the general population might be different if a more representative general population sample had been used in the study. Fourth, the sample size for the study was powered enough for overall comparison (mean utility of all heath states) between the general

population and a patient population. It was not powered for comparison between the populations for mild, moderate, and severe health states separately. Therefore, our study findings should be considered exploratory and need to be confirmed with a bigger study. Furthermore, due to limited number of health states and sample size, it was not possible to evaluate impact of individual dimensions of EQ-5D-5L on differences in health state preferences. And, finally, the findings reported here are limited to valuation of hypothetical health states. They may not be generalized to valuation of experienced health states.

Conclusions

Patients with chronic diseases may differ from members of the general population in the strength of their preferences for hypothetical health states. The difference may depend on type of the disease, age of patients and severity of the health states being valued. As a result, using utility values derived from the general population may under-estimate the comparative effectiveness of healthcare interventions for certain type of diseases, such as heart diseases. Larger studies involving more diverse patient groups and a wider range of health states would be required to confirm our study findings.

Acknowledgments

We acknowledge the assistance of Grace Yang and Qu Limin (Debra) in collecting cancer patients' data. We thank Yeoh Yen Shing for the assistance in data management. The authors appreciate the support of Duke-NUS/SingHealth Academic Research Institute and the medical editing assistance of Serene Ong (Medical writer, Duke-NUS Medical School). The Clinical Trials Research Office, National Heart Centre Singapore has provided administrative support.

Compliance with ethical standards

Funding This study is partially supported by the EuroQol Research Foundation.

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

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

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

References

1. Brazier, J. E., Dixon, S., Ratcliffe, J. (2009). The role of patient preferences in cost-effectiveness analysis. Pharmacoeconomics, 27, 9, 705-12.

2. Stamuli, E. (2011). Health outcomes in economic evaluation: who should value health? British medical bulletin, 97, 197-210.

3. Stiggelbout, A. M., Vogel-Voogt, D. (2008). Health state utilities: a framework for studying the gap between the imagined and the real. Value in Health, 11, 1, 76-87.

4. Drummond, M. F., McGuire, A. (2001). Economic evaluation in health care: merging theory with practice. USA: Oxford University Press.

5. Gandjour, A. (2010). Theoretical foundation of patient v. population preferences in calculating QALYs. Medical Decision Making, 30, 4, E57-63.

6. Wang, P., Tai, E., Thumboo, J., Vrijhoef, H. J. Luo, N. (2014). Does diabetes have an impact on health-state utility? A study of asians in singapore. The Patient, 7, 3, 329-37.

7. Rowen, D., Mulhern, B., Banerjee, S., Tait, R., Watchurst, C., Smith, S. C., et al. (2014). Comparison of general population, patient, and carer utility values for dementia health States. Medical Decision Making, 35, 1, 68-80.

8. Peeters, Y., Stiggelbout, A. M. (2010). Health State Valuations of Patients and the General Public Analytically Compared: A Meta-Analytical Comparison of Patient and Population Health State Utilities. Value in Health,13, 2, 306-09.

9. Gandhi, M., Thumboo, J., Luo, N., Wee, H. L, Cheung, Y. B. (2015). Do chronic disease patients value generic health states differently from individuals with no chronic disease? A case of a multicultural Asian population. Health and quality of life outcomes, 13, 8.

10. Gerhards, S. A., Evers, S. M., Sabel, P. W., Huibers, M. J. (2011). Discrepancy in rating health-related quality of life of depression between patient and general population. Quality of Life Research, 20, 2, 273-79.

11. Dolders, M. G., Zeegers, M. P., Groot, W., Ament, A. (2006). A meta-analysis demonstrates no significant differences between patient and population preferences. Journal of Clinical Epidemiology, 59, 7, 653-64.

12. Mulhern, B., Rowen, D., Snape, D., Jacoby, A., Marson, T., Hughes, D., et al. (2014). Valuations of epilepsy-specific health states: a comparison of patients with epilepsy and the general population. Epilepsy & Behavior, 36, 12-17.

13. Pickard, A. S., Tawk, R., Shaw, J. W. (2013). The effect of chronic conditions on stated preferences for health. The European Journal of Health Economics, 14, 4, 697-702.

14. Krabbe, P. F., Tromp, N., Ruers, T. J., Reil, P. L. (2011). Are patients' judgments of health status really different from the general population? Health and quality of life outcomes, 9, 31.

15. Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., et al. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Quality of life research, 20, 10, 1727-36.

16. Pickard, A. S., De Leon, M. C., Kohlmann, T., Cella, D., Rosenbloom, S. (2007). Psychometric comparison of the standard EQ-5D to a 5 level version in cancer patients. Medical care, 45, 3, 259-63.

17. Janssen, M., Pickard, A. S., Golicki, D., Gudex, C., Niewada, M., Scalone, L., et al. (2013). Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Quality of Life Research, 22, 7, 1717-27.

18. Deaton, C., Froelicher, E. S., Wu, L. H., Ho, C., Shishani, K., Jaarsma, T. (2011). The global burden of cardiovascular disease. European Journal of Cardiovascular Nursing, 10, S5-S13.

19. Fitzmaurice, C., Dicker, D., Pain, A., Hamavid, H., Moradi-Lakeh, M., Maclntyre, M. F., et al. (2015). The global burden of cancer 2013. JAMA oncology, 1, 4, 505-27.

20. National Institute for Health and Clinical Excellence. (2013). Guide to the methods of technology appraisal 2013. National Institute for Health and Clinical Excellence. https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781. Accessed 05 March 2017.

21. Luo, N., Li, M., Stolk, E. A., Devlin, N. J. (2013). The effects of lead time and visual aids in TTO valuation: a study of the EQ-VT framework. The European Journal of Health Economics, 14 Suppl 1, S15-24.

22. Oppe, M., Rand-Hendriksen, K., Shah, K., Ramos-Goni, J. M., Luo, N. (2016). EuroQol Protocols for Time Trade-Off Valuation of Health Outcomes. PharmacoEconomics, 34, 10, 993-1004.

23. Oppe, M., Devlin, N. J., van Hout, B., Krabbe, P. F., de Charro, F. (2014). A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value in Health, 17, 4, 445-53.

24. Sonnenberg, F., Beck, J. (1993). Markov models in medical decision making: a practical guide. Medical Decision Making, 4, 322-338.

25. Williams, R. L. (2000). A note on robust variance estimation for cluster-correlated data. Biometrics, 56, 2, 645-46.

26. Nolan, C. M., Longworth, L., Lord, J., Canavan, J. L., Jones, S. E., Kon, S. S., et al. (2016). The EQ-5D-5L health status questionnaire in COPD: validity, responsiveness and minimum important difference. Thorax, 71, 6, 493-500.

27. Department of Statistics Singapore. Singapore Census of Population 2010. http://www.singstat.gov.sg/publications/publications-and-papers/population#census_of_population. Accessed on 02 June 2017.

28. Mann, R., Brazier, J., Tsuchiya, A. (2009). A comparison of patient and general population weightings of EQ-5D dimensions. Health Economics, 18, 3, 363-372.

29. Rupel, V. P., Marko, M. O. (2016). Should Patients Estimate Health Related Quality Of Life? The Results Of Empirical Study. Value in Health, 19, A480.

30. Papageorgiou, K., Vermeulen, K., Schroevers, M., Stiggelbout, A., Buskens, E. Krabbe, P., et al. (2015). Do individuals with and without depression value depression differently? And if so, why?. Quality of Life Research, 24, 11, 2565-2575.

31. Sayah, F., Bansback, N., Bryan, S., Ohinmaa, A., Poissant, L., Pullenayegum, E., et al. (2016). Determinants of time trade-off valuations for EQ-5D-5L health states: data from the Canadian EQ-5D-5L valuation study. Quality of Life Research, 25, 7, 1679-1685.

32. Shaw, J., Johnson, J., Chen, S., Levin, J., Coons, S. (2007). Racial/ethnic differences in preferences for the EQ-5D health states: results from the U.S. valuation study. Journal of Clinical Epidemiology, 60, 6, 479-490.

Tables

Table 1 Demographic and health characteristics of the study participants

Characteristic	General population (N=169)	Heart disease patients (N=157)		Cancer patients (N=169)	
	n (%)	n (%)	P-value*	n (%)	P-value*
Female	89 (52.7)	54 (34.4)	0.001	98 (58.0)	0.381
Age (years)			<0.001		<0.001
21-40	75 (44.4)	13 (8.3)		20 (11.8)	
41-60	65 (38.5)	61 (38.9)		93 (55.0)	
>60	29 (17.2)	83 (52.9)		56 (33.1)	
Mean (SD)#	44.4 (14.2)	58.8 (11.6)	<0.001	54.4 (13.0)	<0.001
Ethnicity			0.029		0.079
Chinese	140 (82.8)	110 (70.1)		123 (72.8)	
Malay	10 (5.9)	12 (7.6)		22(13.0)	
Indian	10 (5.9)	23 (14.7)		10 (5.9)	
Others	9 (5.3)	12 (7.6)		14 (8.3)	
Education level			<0.001		<0.001
Primary (6 years) or less	12 (7.1)	35 (22.3)		25 (14.8)	
Secondary (7-11 years)	54 (32.0)	88 (56.1)		91 (53.9)	
Diploma, University or higher	103 (61.0)	34 (21.7)		53 (31.4)	
Married	104 (61.5)	101 (64.3)	0.647	113 (66.9)	0.364
Employed	109 (64.5)	84 (53.5)	0.055	95 (56.2)	0.148
Household earnings per month			<0.001		0.010
<S$4000	60 (35.5)	98 (62.4)		87 (51.5)	
>=S$4000	91 (53.9)	44 (28.0)		65 (38.5)	
Don't know/refused	18 (10.7)	15 (9.6)		17 (10.1)	
Religion			0.693		<0.001
No religious belief	37 (21.9)	26 (16.6)		15 (8.9)	
Buddhism/Taoism	55 (32.5)	54 (34.4)		66 (39.1)	
Christians	44 (26.0)	43 (27.4)		46 (27.2)	
Islam	15 (8.9)	19 (12.1)		34 (20.1)	
Others	18 (10.7)	15 (9.6)		8 (4.7)	
Self-reported health on VAS#	82.6 (10.4)	73.8 (14.5)	<0.001	76.3 (17.3)	<0.001
Self-reported chronic diseases					
Cancer	1 (0.6)	8 (5.1)	0.016	169 (100)	<0.001
Heart disease	6 (3.6)	157 (100)	<0.001	6 (3.6)	1.000
Hypertension	25 (14.8)	90 (57.3)	<0.001	48 (28.4)	0.003
Hyper/dyslipidemia	20 (11.8)	92 (58.6)	<0.001	29 (17.2)	0.216
Diabetes mellitus	7 (4.1)	46 (29.3)	<0.001	33 (19.5)	<0.001
Arthritis/gout/joint pain	8 (4.7)	23 (14.7)	0.002	11 (6.5)	0.638
Lung disease	16 (9.5)	12 (7.6)	0.693	7 (4.1)	0.082
Other	15 (8.9)	37 (23.6)	<0.001	19 (11.2)	0.588
Number of self-reported chronic diseases					
None	107 (63.3)	0 (0.0)	-	0 (0.0)	-

Characteristic	General population (N=169)	Heart disease patients (N=157)		Cancer patients (N=169)	
	n (%)	n (%)	P-value*	n (%)	P-value*
One	38 (22.5)	24 (15.3)	-	84 (49.7)	-
More than one	24 (14.2)	133 (84.7)	-	85 (50.3)	-

* Comparison with the general population group using Fisher's exact test or two-sample independent t test

Mean (Standard deviation)

VAS: Visual Analogue Scale (100: Best imaginable health state, 0: Worst imaginable health state)

Table 2 Comparison of composite time trade-off utility values between the general and patient populations

EQ-5D-5L heath states	General population (N=169)	Heart disease patients (N=157)	Cancer patients (N=169)
All health states			
Mean (SD)	0.398 (0.657)	0.291 (0.807)	0.336 (0.723)
Mean difference (95% CI)[1,4]		-0.107 (-0.191, -0.022)*	-0.062 (-0.147, 0.022)
Adjusted mean difference (95% CI)[1,5]		0.031 (-0.067, 0.130)	-0.007 (-0.058, 0.045)
Mild health states[6]			
Mean (SD)	0.842 (0.255)	0.875 (0.345)	0.829 (0.362)
Mean difference (95% CI)[1,4]		0.033 (-0.018, 0.084)	-0.013 (-0.067, 0.041)
Adjusted mean difference (95% CI)[1,5]		0.076 (0.021, 0.131)**	-0.009 (-0.072, 0.054)
Moderate health states[6]			
Mean (SD)	0.505 (0.545)	0.452 (0.704)	0.417 (0.651)
Mean difference (95% CI)[1,4]		-0.053 (-0.166, 0.059)	-0.088 (-0.196, 0.020)
Adjusted mean difference (95% CI)[1,5]		0.091 (-0.041, 0.223)	-0.011 (-0.132, 0.111)
Severe health states[6]			
Mean (SD)	-0.016 (0.693)	-0.267 (0.764)	-0.096 (0.719)
Mean difference (95% CI)[1,4]		-0.251 (-0.380, -0.123)**	-0.080 (-0.206, 0.045)
Adjusted mean difference (95% CI)[1,5]		-0.048 (-0.198, 0.103)	-0.002 (-0.137, 0.133)
11122			
Mean (SD)	0.889 (0.189)	0.903 (0.274)	0.879 (0.287)
Mean difference (95% CI)[1,2]		0.013 (-0.038, 0.065)	-0.010 (-0.062, 0.042)
Adjusted mean difference (95% CI)[1,3]		0.031 (-0.030, 0.093)	-0.020 (-0.079, 0.038)
21121			
Mean (SD)	0.879 (0.204)	0.925 (0.256)	0.871 (0.290)
Mean difference (95% CI)[1,2]		0.046 (-0.004, 0.096)	-0.008 (-0.062, 0.045)
Adjusted mean difference (95% CI)[1,3]		0.067 (0.007, 0.128)*	-0.015 (-0.076, 0.046)
21222			
Mean (SD)	0.759 (0.329)	0.798 (0.457)	0.738 (0.464)
Mean difference (95% CI)[1,2]		0.040 (-0.047, 0.126)	-0.021 (-0.107, 0.065)
Adjusted mean difference (95% CI)[1,3]		0.129 (0.025, 0.232)*	0.009 (-0.088, 0.106)
21232			
Mean (SD)	0.640 (0.436)	0.681 (0.541)	0.592 (0.557)
Mean difference (95% CI)[1,2]		0.041 (-0.066, 0.147)	-0.049 (-0.156, 0.059)
Adjusted mean difference (95% CI)[1,3]		0.148 (0.020, 0.276)*	-0.005 (-0.124, 0.115)
32232			
Mean (SD)	0.532 (0.523)	0.393 (0.724)	0.433 (0.623)
Mean difference (95% CI)[1,2]		-0.138 (-0.275, -0.001)*	-0.098 (-0.221, 0.025)
Adjusted mean difference (95% CI)[1,3]		-0.001 (-0.162, 0.160)	-0.018 (-0.155, 0.119)
32333			
Mean (SD)	0.343 (0.622)	0.282 (0.768)	0.227 (0.714)
Mean difference (95% CI)[1,2]		-0.062 (-0.214, 0.090)	-0.117 (-0.260, 0.026)
Adjusted mean difference (95% CI)[1,3]		0.127 (-0.051, 0.304)	-0.009 (-0.165, 0.147)
31242			
Mean (SD)	0.226 (0.648)	0.035 (0.785)	0.109 (0.726)
Mean difference (95% CI)[1,2]		-0.191 (-0.347, -0.034)*	-0.117 (-0.264, 0.031)

EQ-5D-5L heath states	General population (N=169)	Heart disease patients (N=157)	Cancer patients (N=169)
Adjusted mean difference (95% CI)[1,3]		-0.070 (-0.247, 0.107)	-0.060 (-0.222, 0.101)
22224			
Mean (SD)	0.069 (0.703)	-0.069 (0.825)	0.101 (0.742)
Mean difference (95% CI)[1,2]		-0.138 (-0.304, 0.029)	0.032 (-0.123, 0.186)
Adjusted mean difference (95% CI)[1,3]		0.086 (-0.103, 0.274)	0.148 (-0.023, 0.319)
53343			
Mean (SD)	-0.165 (0.672)	-0.545 (0.614)	-0.309 (0.645)
Mean difference (95% CI)[1,2]		-0.380 (-0.520, -0.239)**	-0.144 (-0.285, -0.003)*
Adjusted mean difference (95% CI)[1,3]		-0.141 (-0.298, 0.016)	-0.064 (-0.217, 0.090)
33453			
Mean (SD)	-0.193 (0.665)	-0.490 (0.641)	-0.284 (0.649)
Mean difference (95% CI)[1,2]		-0.297 (-0.440, -0.155)**	-0.091 (-0.232, 0.049)
Adjusted mean difference (95% CI)[1,3]		-0.065 (-0.226, 0.096)	-0.032 (-0.185, 0.122)

[1] Difference: mean utility of the patients group minus mean utility of the general population group

[2] Using ordinary least-square regression model.

[3] Using ordinary least square regression model adjusted for gender, age, ethnicity, education level, employment status, household earnings, and religion.

[4] Using ordinary least-square regression model (with robust standard error for cluster data).

[5] Using ordinary least square regression model (with robust standard error for cluster data) adjusted for gender, age, ethnicity, education level, employment status, household earnings, and religion.

[6] EQ-5D-5L health states with all dimensions at severity level either 1 or 2 are considered ‘mild’ heath states. EQ-5D-5L health states with at least one dimension at severity level either 4 or 5 are considered ‘severe’ heath states. EQ-5D-5L health states which are neither ‘mild’ or ‘severe’ are considered ‘moderate’ health states.

* P-value <0.05; ** P-value <0.01

SD: Standard deviation; CI: Confidence interval

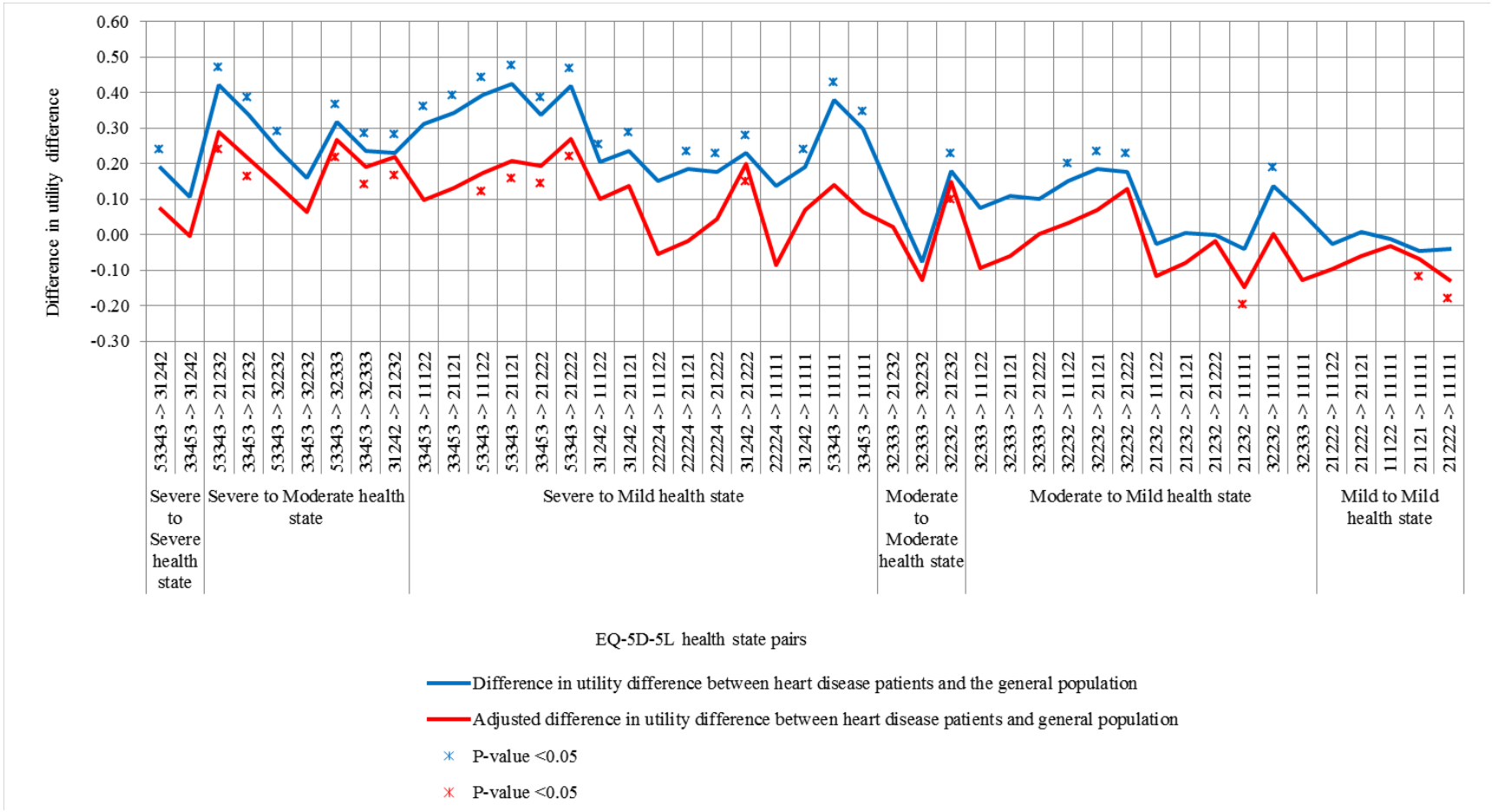
Figure Captions

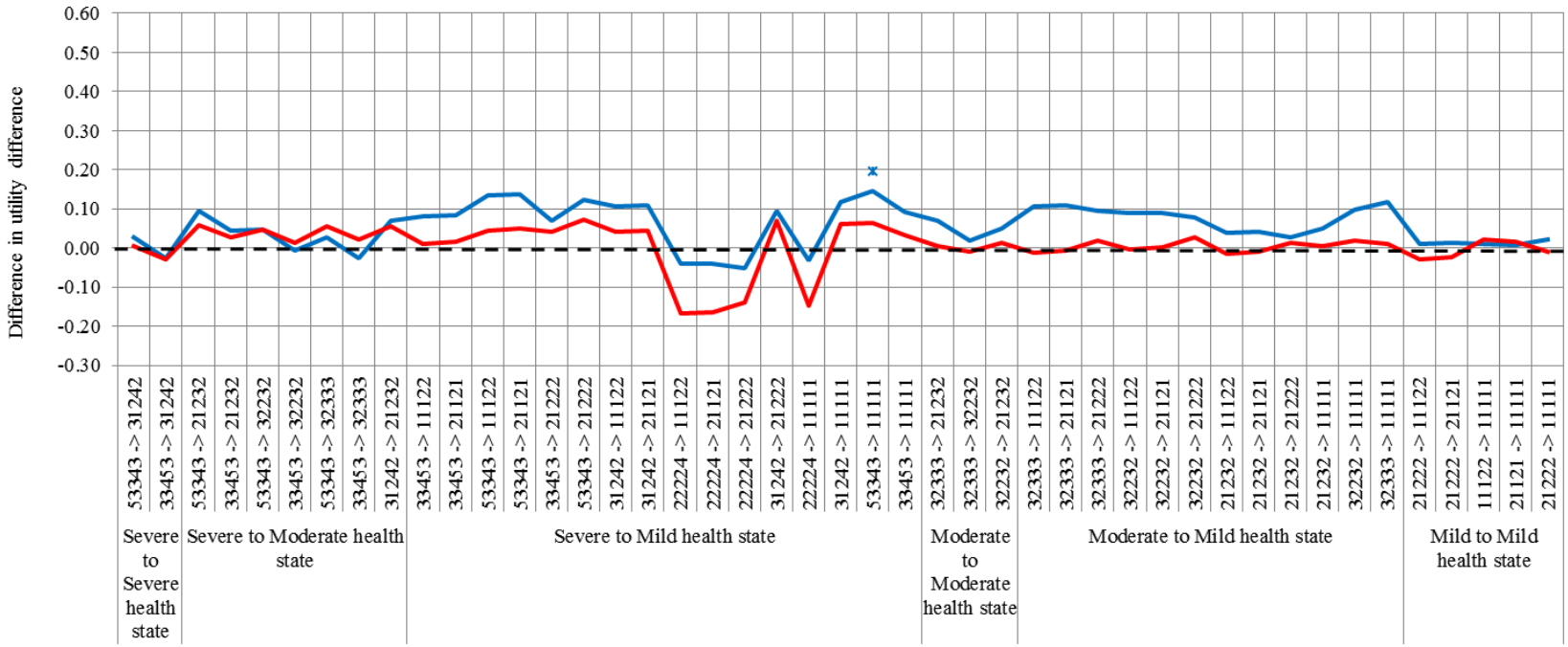
Fig. 1 Estimated differences in differences in utility values between heart disease patients and the general population for pair of health states

Difference in utility values defined as utility value of a better health state minus that of a poorer health state.

Fig. 2 Estimated differences in differences in utility values between cancer patients and the general population for pair of health states

Difference in utility values defined as utility value of a better health state minus that of a poorer health state.


Difference in utility difference
0.60
0.50
0.40
0.30
0.20
0.10
0.00
-0.10
-0.20
-0.30
53343 -> 31242
33453 -> 31242
53343 -> 21232
33453 -> 21232
53343 -> 32232
33453 -> 32232
53343 -> 32333
33453 -> 32333
31242 -> 21232
33453 -> 11122
33453 -> 21121
53343 -> 11122
53343 -> 21121
33453 -> 21222
53343 -> 21222
31242 -> 11122
31242 -> 21121
22224 -> 11122
22224 -> 21121
22224 -> 21222
31242 -> 21222
22224 -> 11111
31242 -> 11111
53343 -> 11111
33453 -> 11111
32333 -> 21232
32333 -> 32232
32232 -> 21232
32333 -> 11122
32333 -> 21121
32333 -> 21222
32232 -> 11122
32232 -> 21121
32232 -> 21222
21232 -> 11122
21232 -> 21121
21232 -> 21222
21232 -> 11111
32232 -> 11111
32333 -> 11111
21222 -> 11122
21222 -> 21121
11122 -> 11111
21121 -> 11111
21222 -> 11111
Severe to Severe health state
Severe to Moderate health state
Severe to Mild health state
Moderate to Moderate health state
Moderate to Mild health state
Mild to Mild health state
EQ-5D-5L health state pairs
Difference in utility difference between heart disease patients and the general population
Adjusted difference in utility difference between heart disease patients and general population
P-value <0.05
P-value <0.05



EQ-5D-5L health state pairs

Difference in utility difference between cancer patients and the general population

Adjusted difference in utility difference between cancer patients and general population

* P-value <0.05

No difference in utility difference was statistically significant at 5% level.

Electronic Supplementary Material

Article title: Comparison of Health State Values Derived from Patients and Individuals from the General Population

Journal: Quality of Life Research

Authors: Mihir Gandhi, Ru San Tan, Raymond Ng, Su Pin Choo, Whay Kuang Chia, Chee Keong Toh, Carolyn Lam, Phong Teck Lee, Nang Khaing Zar Latt, Kim Rand-Hendriksen, Yin Bun Cheung, Nan Luo

Corresponding author: Nan Luo; ephln@nus.edu.sg

Appendix Table 1 Comparison of demographic and health characteristics of participants included and excluded from the analysis

Characteristic, n (%)	Included participants (N = 495)	Excluded participants (N = 30)	P-value*
Population			0.011
General population	169 (34.1)	6 (20.0)	
Heart disease patients	157 (31.7)	18 (60.0)	
Cancer patients	169 (34.1)	6 (20.0)	
Female	241 (48.7)	15 (50.0)	1.000
Age (years)			0.198
21-40	108 (21.8)	3 (10.0)	
41-60	219 (44.2)	13 (43.3)	
>60	168 (33.9)	14 (46.7)	
Mean (SD)#	52.4 (14.3)	56.8 (12.1)	0.099
Ethnicity			0.516
Chinese	373 (75.4)	21 (70.0)	
Malay	44 (8.9)	2 (6.7)	
Indian	43 (8.7)	5 (16.7)	
Others	35 (7.1)	2 (6.7)	
Education level			0.134
Primary (6 years) or less	72 (14.6)	3 (10.0)	
Secondary (7-11 years)	233 (47.1)	20 (66.7)	
Diploma, University or higher	190 (38.4)	7 (23.3)	
Married	318 (64.2)	23 (76.7)	0.236
Employed	288 (58.2)	17 (56.7)	0.507
Household earnings per month			0.820
<S$4000	245 (49.5)	17 (56.7)	
>=S$4000	200 (40.4)	11 (36.7)	
Don’t know/refused	50 (10.1)	2 (6.7)	
Religion			0.416
No religious belief	78 (15.8)	2 (6.7)	
Buddhism/Taoism	175 (35.4)	10 (33.3)	
Christians	133 (26.9)	9 (30.0)	
Islam	68 (13.7)	4 (13.3)	
Others	41 (8.3)	5 (16.7)	
Self-reported health on VAS#	77.7 (14.8)	75.6 (16.6)	0.467

* Comparison with the general population group using Fisher’s exact test or two-sample independent t test

Mean (Standard deviation)

VAS: Visual Analogue Scale (100: Best imaginable health state, 0: Worst imaginable health state)

Appendix Table 2 Comparison of demographic characteristics of the study sample with the Singaporean general population according to Census 2010.

	Census 2010	Study sample		
Characteristic, % of the population	Singapore general population of age 20 – 79 years	General population (N=169)	Heart disease patients (N=157)	Cancer patients (N=169)
Female	50.9	52.7	34.4	58.0
Age (years)				
20-39	40.9	41.4	7.6	11.8
40-59	42.5	39.6	36.9	51.5
≥60	16.6	18.9	55.4	36.7
Ethnicity				
Chinese	75.8	82.8	70.1	72.8
Malay	12.1	5.9	7.6	13.0
Indian	8.8	5.9	14.7	5.9
Others	3.2	5.3	7.6	8.3
Education level*				
Primary (6 years) or less	8.2	7.1	22.3	14.8
Secondary (7-11 years)	34.2	32.0	56.1	53.9
Diploma, University or higher	57.6	61.0	21.7	31.4
Married	65.8	61.5	64.3	66.9
Household earnings per month				
<S$4000	37.5	35.5	62.4	51.5
>=S$4000	62.5	53.9	28.0	38.5
Don’t know/refused	-	10.7	9.6	10.1
Religion				
No religious belief	16.8	21.9	16.6	8.9
Buddhism/Taoism	44.8	32.5	34.4	39.1
Christians	18.3	26.0	27.4	27.2
Islam	14.2	8.9	12.1	20.1
Others	5.9	10.7	9.6	4.7

* Excluded population with no educational qualification.

Appendix Table 3 Disease characteristics of heart disease patients

Characteristic, n (%)	Heart disease patients (N=157)
Heart disease type	
Coronary artery disease	99 (63.1)
Arrhythmias	42 (26.8)
Heart failure	31 (19.8)
Other	7 (4.5)
NYHA functional classification*	
I	112 (71.3)
II	37 (23.6)
III	8 (5.1)
CCS functional classification*	
I	114 (72.6)
II	22 (14.0)
III	17 (10.8)
IV	4 (2.6)
Years from most recent episode of symptom onset	
Mean (SD)	1.4 (2.0)
Heart disease treatment received	
Medicinal therapy	157 (100.0)
Surgery	115 (73.3)
Cardiac rehabilitation	45 (28.7)
Lifestyle changes	22 (14.0)

NYHA: New York Heart Association; CCS: Canadian Cardiovascular Society; SD: Standard Deviation

* All subjects were graded on symptoms of breathlessness and chest pain using these scores regardless of whether they had known heart failure or coronary artery disease, respectively.

Appendix Table 4 Disease characteristics of cancer patients

Characteristic, n (%)	Cancer patients (N=169)
Cancer type	
Breast	49 (29.0)
Colorectal	36 (21.3)
Lymphoma	19 (11.2)
Lung	13 (7.7)
Prostate	9 (5.3)
Ovary	7 (4.1)
Stomach	6 (3.6)
Nasopharyngeal	6 (3.6)
Other	32 (18.9)
Highest cancer stage	
0	9 (5.3)
1	26 (15.4)
2	31 (18.3)
3	41 (24.3)
4	62 (36.7)
Years from diagnosis of the most recent cancer	
Mean (SD)	1.9 (1.4)
ECOG performance status	
0	95 (56.2)
1	73 (43.2)
2	1 (0.6)
Cancer treatment received	
Chemotherapy	124 (73.4)
Surgery	113 (66.9)
Radiation therapy	71 (42.0)
Hormone therapy	33 (19.5)
Immune therapy	21 (12.4)
Targeted therapy	13 (7.7)
Other therapy	13 (7.7)

ECOG: Eastern Cooperative Oncology Group

Appendix Table 5 Summary of regression model (adjusted for covariates) for comparing utility values of all health states between patients and the general populations

Variables	Comparison of utility values between heart disease patients and the general population		Comparison of utility values between cancer patients and the general population	
	Coefficient (95% CI)	P-value	Coefficient (95% CI)	P-value
Population				
General	Reference		Reference	
Patients	0.031 (-0.067, 0.130)	0.535	-0.007 (-0.058, 0.045)	0.803
Gender				
Male	Reference		Reference	
Female	-0.073 (-0.156, 0.011)	0.089	-0.028 (-0.077, 0.020)	0.253
Age (years)		[<0.0001]		[<0.0001]
21- 40	Reference		Reference	
41- 60	-0.147 (-0.247, -0.08)	0.004	-0.091 (-0.155, -0.028)	0.005
> 60	-0.288 (-0.415, -0.161)	<0.001	-0.220 (-0.300, -0.141)	<0.001
Ethnicity		[0.457]		[0.232]
Chinese	Reference		Reference	
Malay	0.100 (-0.133, 0.334)	0.398	-0.056 (-0.212, 0.100)	0.482
Indian	-0.050 (-0.238, 0.138)	0.600	0.101 (-0.040, 0.242)	0.159
Others	-0.050 (-0.250, 0.150)	0.622	-0.024 (-0.142, 0.095)	0.697
Education level		[0.053]		[0.054]
Primary (6 years) or less	Reference		Reference	
Secondary (11 years)	0.084 (-0.043, 0.211)	0.192	0.071 (-0.009, 0.151)	0.082
Diploma, university or higher	0.186 (0.031, 0.341)	0.019	0.117 (0.022, 0.212)	0.016
Marital status				
Married	Reference		Reference	
Single	-0.068 (-0.154, 0.017)	0.117	-0.061 (-0.111, -0.010)	0.018
Employment status				
Unemployed	Reference		Reference	
Employed	-0.042 (-0.130, 0.046)	0.346	-0.025 (-0.078, 0.028)	0.362
Household earning per month		[0.771]		[0.002]
< S$4000	Reference		Reference	
⩾ S$ 4000	0.023 (-0.077, 0.123)	0.647	0.092 (0.035, 0.149)	0.002
Don't know/refused	-0.027 (-0.169, 0.114)	0.704	-0.024 (-0.108, 0.059)	0.566
Religion		[0.037]		[0.001]
No religious belief	Reference		Reference	
Buddhism/Taoism	0.130 (0.012, 0.247)	0.030	0.111 (0.037, 0.184)	0.003
Christians	0.160 (0.030, 0.290)	0.016	0.139 (0.063, 0.215)	<0.001
Islam	0.100 (-0.135, 0.334)	0.404	0.163 (0.009, 0.317)	0.038
Others	-0.028 (-0.217, 0.161)	0.769	-0.022 (-0.149, 0.106)	0.740
Constant	0.356 (0.155, 0.557)	0.001	0.299 (0.177, 0.421)	<0.001

CI: Confidence Interval. P-values in square brackets were calculated for joint tests of the categorical variables using the Wald test.

PUBLICATION IV

Sample size determination for EQ-5D-5L value set studies

Mihir Gandhi, Ying Xu, Nan Luo, Yin Bun Cheung

Quality of Life Research 2017, 26(12):3365-3376
https://doi.org/10.1007/s11136-017-1685-3

Sample size determination for EQ-5D-5L value set studies

Mihir GANDHI[a,b,c]*, Ying XU[b], Nan LUO[e], Yin Bun CHEUNG[b,c]

[a]Department of Biostatistics, Singapore Clinical Research Institute, #02-01, 31 Biopolis Way, Singapore

[b]Centre for Quantitative Medicine, Duke-NUS Medical School, Level 6, Academia, 20 College Road, Singapore

[c]Tampere Center for Child Health Research, University of Tampere and Tampere University Hospital, Arvo Building, Lääkärinkatu 1, Tampere, Finland

[e]Saw Swee Hock School of Public Health, National University of Singapore, Tahir Foundation Building (MD1), 12 Science Drive 2, Singapore

Email addresses of co-authors:

Ying Xu: tinayxu@gmail.com (ORCID: 0000-0001-7712-6283)

Nan Luo: ephln@nus.edu.sg (ORCID: 0000-0001-7980-6979)

Yin Bun Cheung: yinbun.cheung@duke-nus.edu.sg (ORCID: 0000-0003-0517-7625)

*Corresponding author:

Mihir Gandhi

Head, Department of Biostatistics

Singapore Clinical Research Institute, #02-01, Nanos, 31 Biopolis Way, Singapore.

Email: mihir.gandhi@scri.edu.sg

Telephone: +65-65088309

ORCID: 0000-0002-8902-2710

Acknowledgements

We thank Juan Manuel Ramos Goñi (Senior Scientist, The EuroQol Office) and Lu Qingshu (Senior Biostatistician, Singapore Clinical Research Institute) for helpful comments on earlier versions of this article. The authors appreciate the support of Duke-NUS/SingHealth Academic Research Institute and the medical editing assistance of Serene Ong (Medical writer, Duke-NUS Medical School).

Purpose: The EuroQol 5-Dimension (EQ-5D) is a widely-used health status instrument for cost-utility analysis of healthcare interventions. Recently, its 5-Level version (EQ-5D-5L) and a protocol for conducting valuation of its health states were developed. We propose four approaches for estimating the sample size for EQ-5D-5L valuation according to the standardized procedures of the protocol.

Methods: The first approach is for estimating mean health state utility values with a desired precision level using a regression model. The second approach, empirical in nature, determines a sample size based on mean absolute error in predicting health state values using a large-scale reference study. The last two approaches are for assessing the significance of regression coefficients of health state descriptors and to estimate the regression coefficients with a desired precision for predicting health state utility values.

Results: Using data from a Singaporean study, we estimated parameters that are useful for sample size determination, including the design effect. Each of the approaches was illustrated with examples and pragmatic recommendations were provided.

Conclusions: Capitalizing on the EQ-5D-5L valuation protocol, we proposed four sample size estimation approaches which can help to decide an appropriate sample size for a value set study.

Introduction

The 3-Level EuroQol 5-Dimension (EQ-5D) is a widely used health-status instrument for cost-utility analysis of healthcare interventions [1]. Recently, its 5-Level version (EQ-5D-5L) was developed, thereby increasing fineness in health-state descriptors and resulting in improved accuracy of measurement [2]. It requires value sets for use in economic evaluations. Previous EQ-5D value set studies were conducted using a variety of valuation methods, such as standard gamble, time trade-off (TTO), visual analogue scale. Although, TTO has been widely used, it was implemented in numerous ways and the value sets developed were not comparable across studies [3]. Learning from past experiences and performing pilot studies for optimizing valuation methodology, the EuroQol Group has developed a new valuation protocol for the EQ-5D-5L, namely the EuroQol Valuation Technology (EQ-VT) protocol [4]. This protocol is currently considered as a 'gold standard' for developing EQ-5D-5L value sets. The protocol has been adopted by many countries including England, Japan, Canada, Netherlands, Korea and Singapore in developing their country-specific value sets in a standard manner [5].

The protocol recommends a sample size of 1000 members of the general public [4]. This recommendation for the sample size, however, is based on some assumptions without support from empirical data and provided limited theoretical justification. On one hand the EQ-VT protocol addresses the issue of comparability of valuation methods, the suggested sample size could be a limiting factor for many countries where resources for conducting such studies is a major constrain. So far there is no follow-up evaluation on whether the recommended sample size is sufficient or excessive for developing a value set. The selection of an optimal sample size is very important from both statistical and feasibility points of view. This article aims to propose four sample size estimation methods for an EQ-5D-5L value set study. These sample size methods are illustrated using input parameters estimated from a recently completed EQ-5D-5L value set study in Singapore. This article will help guide a statistically sound decision on sample size estimation for future value set studies, in order to achieve optimal utilization of resources and confidence with the value set results.

The plan for the article is as follows. In the Methods section, first we described the EQ-VT protocol and current practice for the data analysis. Next, based on the EQ-VT protocol, we derived mathematical formula for estimating several parameters, which are critical for sample size determination. The Results section provides other key parameters that are needed for the use of the proposed methods, estimated from a large-scale study in Singapore. It also illustrates the use of the methods. The article ends with Discussion followed by Conclusions section.

The EQ-VT protocol uses the composite Time Trade-Off (cTTO) for valuation. It also tests the use of discrete-choice experiment as a potential method [6]. This article focuses on sample size estimation for the cTTO based value sets.

Methods

The EQ-VT protocol

The EQ-5D is a generic health status instrument [7]. It contains five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), and a visual analogue scale (EQ-VAS) of the overall health status. The new version of EQ-5D, EQ-5D-5L, describes each dimension with five levels of severity (broadly corresponding to no problem, slight problems, moderate problems, severe problems, and extreme problems). Thus, it can describe 3125 possible health states.

According to the EQ-VT protocol, each participant values a randomly selected set (called a *block*) of 10 hypothetical EQ-5D-5L health states [4]. Each block involves one very mild health state from five pre-specified ones (21111, 12111, 11211, 11121, 11112), the most severe health state (55555), and eight health states from 80 pre-specified health states (from the remaining 3119 possible health states). Here, the health state '21111' indicates slight problems (level 2 severity) in the first dimension - mobility, no problem (level 1 severity) in the remaining four dimensions. Similarly, the other health states are defined. The protocol contains a total of 10 unique blocks (Table 1). Simple random sampling is used to select one of the 10 blocks for each participant; hence equal probability of each block being chosen.

The cTTO method requires participants to identify the point of preferential indifference between 10 years of life in the described target health state, followed by death, and a shorter life ($x \leq 10$ years) in full health, followed by death. With a defined utility value of 1 for 10 years in full health, the utility value of the target health state can be calculated as $x/10$. For health states considered to be worse than death, a lead-time of 10 years is added to both alternatives in order to elicit a negative utility value for the health state. The utility value of a worse than death health state is calculated as $(x-10)/10$ such that the utility value of each health state is bounded at -1 and 1; 0 represents value for the 'dead' state. More details on the EQ-VT protocol are published elsewhere [4, 6].

Development of a value set

A value set provides an algorithm for deriving a utility value for each health state. A value set study conducted using the EQ-VT protocol provides values for only 86 directly elicited health states. These values are used to estimate values for all 3125 health states of the EQ-5D-5L. There are potentially various ways to achieve it using an appropriate statistical model such as an ordinary least squares (OLS) model, OLS model with cluster-robust standard error (SE), random effects (RE) model, interval regression model, and so on [8]. Currently, there is no consensus on which model is ideal for predicting utility values. Usually, a study team performs several models and decides upon a final model based on the models' performance for consistency, bias, precision, and parsimony.

As an initial choice, an OLS model for estimating utility values can be specified as:

$$y = X\beta + \varepsilon. \quad (1)$$

We call this model as the Basic OLS model. Here, $y = (y_1, y_2, \dots, y_n)'$, where $y_i = (y_{i1}, y_{i2}, \dots, y_{i10})'$ is a vector of utility values of 10 health states elicited from the i^{th} participant, and n is the number of participants; $\beta = (\beta_0, \beta_1, \dots, \beta_{20})'$ is a vector of regression coefficients to be estimated, corresponding to the intercept and 20 indicator variables representing the four severity levels (2 to 5 levels) of the five dimensions (dummy coding scheme); X is an $10n \times 21$ design matrix with the first column being an identity vector for the intercept and the remaining 20 columns for the indicator variables; ε is an $10n \times 1$ vector of errors. It is assumed that ε_{ij} (error for utility value of j^{th} health state in i^{th} participant) ~ i.i.d. $N(0, \sigma^2)$, where i.i.d. stands for independent and identically distributed and $N(0, \sigma^2)$ denotes a normal distribution with mean 0 and variance σ^2.

As each participant values 10 health states, values of health states elicited by the same participant might be correlated. Furthermore, the variance may not be constant. Thus, it is advisable to use cluster-robust SE (considering each participant as a cluster) to make valid statistical inference about the coefficients of the model [9].

Alternatively, one can use a RE model for utility values specified as [10]:

$$y = X\beta + \gamma + \varepsilon. \quad (2)$$

We call this model as the Basic RE model. Here, *y*, *X*, and β are vectors as mentioned in equation (1); $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n)'$, where $\gamma_i = (\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{i10})'$, $\gamma_{ij} = \gamma_{ik}$ for all $j \neq k$, represents a 10 × 1 vector of i^{th} participant-specific random intercept which underlines the intra-participant correlation among the 10 health states; ε is an $10n \times 1$ vector of errors. It is assumed that $\gamma_{ij} \sim N(0, \sigma_\gamma^2)$, $\varepsilon_{ij} \sim N(0, \sigma_e^2)$, and γ_i's and ε_{ij}'s are independent. Here, σ_γ^2 represents the between-participant variance, σ_e^2 is the within-participant variance,$Cov(y_{ij}, y_{ik}) = \sigma_\gamma^2$ for all $j \neq k$, $Cov(y_{ij}, y_{lk}) = 0$ for all $i \neq l$ and $Var(y_{ij}) = \sigma_\gamma^2 + \sigma_e^2 = \sigma^2$. In other words, within-participant utility values are correlated, but between-participant utility values are uncorrelated. Furthermore, intra-participant correlation does not depend on order of health states being valued.

The models in equations (1) and (2) may be extended to include additional variables to indicate whether a health state is mild (e.g., all dimensions at severity level 1 or 2) or severe (e.g., at least one dimension at severity level 4 or 5) and possibly adding interactions between the health state descriptor variables to improve model performance for the country-specific valuation data [11]. Alternative modeling technique such as a quantiles regression should also be considered. Furthermore, indicator variables for health state descriptor may be coded with backward difference coding scheme [12] so that regression coefficients of each dimension represent estimated mean differences in the utility values compared to their previous level of severity. In this article, we have used dummy coding scheme for illustration as it is widely-used. However, relevant information regarding the difference coding scheme is provided in Electronic Supplementary Material.

Prerequisite parameters for sample size estimation

Inverse of cross-product of design matrix

As will be shown in the subsequent sections, the inverse of the cross-product of the design matrix $C = (X'X)^{-1}$, where *X* is as defined in equation (1), plays an important role in the sample size determination. We thus derive this piece of generic information before discussing specific sample size determination methods.

It is intuitive to begin with considering the sample size as a multiple of 10. Since the valuation protocol has only 10 unique blocks of health states (Table 1), each block has 10 health states, and the blocks have equal probability of being selected, for $N_s = 10 \times M$ participants, each of the 10 blocks on average is to be valued *M* times. Thus, for N_s participants, the matrix *X* is expected to be $X = \left(X_{u(1)}, X_{u(2)}, \ldots, X_{u(M)}\right)'$, where $X_{u(i)} = X_u$ is the i^{th} replicate of the 100 × 21 design matrix based on the 10 unique blocks. Therefore,

$$E(C|N_s = 10M) = (X_u'X_u)^{-1}/M = D/M = 10\,D/N_s, \quad (3)$$

where $D = (X_u'X_u)^{-1}$.

As such,

$$E(C_{jj}) = \frac{D_{jj}}{M} = \frac{10D_{jj}}{N_s}, \quad (4)$$

where C_{jj} and D_{jj} are the j^{th} diagonal element of the 21×21 matrices *C* and *D*, respectively.

The matrix *D* for the dummy coding scheme is shown in Table 2, and for the backward difference coding scheme in Electronic Supplementary Material Table A.1.

Design effect

Standard sample size estimation approaches as well as those described in the subsequent sections assume (or begin with assuming) independent observations. However, the EQ-VT protocol requires each participant to value 10 health states. Such clustering of observations within participants impact the variance of the estimates. On the one hand, utility values (outcome variable) elicited by the same participant are likely positively correlated, as some people may generally give higher (or lower) values to health states no matter what these health states are. On the other hand, the EQ-VT protocol is developed to ensure that the 10 health states (exposure variables) are diverse within each participant, leading to negative correlation in the exposure. Given positive correlation in outcomes, the variances of the regression estimates are inflated if the exposures are positively correlated (e.g., cluster randomized trials), or deflated if the exposures are negatively correlated (e.g., crossover drug trials) [13, 14]. To compensate for the variance inflation or deflation, the sample size should be increased (or decreased) by the 'design effect' factor (DE) [13-15]. The sample size adjusted for the design effect is calculated as,

$$N = N_s \times DE\,. \quad (5)$$

For a randomized controlled trial that has one exposure variable, there are formulas for estimation of design effect and therefore sample size using the intra-class correlation coefficients as inputs [13-15]. However, currently there is no such formula for the complex case of multivariable regression analysis. We propose to empirically estimate the design effect using data from a large-scale study, as will be shown in Results section.

The design effect for predicted utility value of a particular health state HS_0 is calculated as

$$DE_{y_0} = \frac{V_{Cluster}(\hat{y}_0)}{V(\hat{y}_0)}, \quad (6)$$

where $V_{Cluster}(\hat{y}_0)$ and $V(\hat{y}_0)$ are cluster-robust and OLS variances, respectively, for the predicted utility value of HS_0. For simplicity, one may consider using simple or weighted mean of DE_{y_0} values ($\overline{DE}_y$) of directly valued health states, where weights can be chosen proportional to probability of occurrence of the health states. According to the valuation protocol, probabilities of occurrences of very mild health states (21111, 12111, 11211, 11121, 11112), the most severe health state (55555) and the remaining 80 directly valued health states are 0.02, 1.0, and 0.01, respectively (Table 1).

Similar to equation (5), the design effect for a regression coefficient, β_j, is calculated as

$$DE_{\beta_j} = \frac{V_{Cluster}(\hat{\beta}_j)}{V(\hat{\beta}_j)}, \quad (7)$$

where $V_{Cluster}(\hat{\beta}_j)$ and $V(\hat{\beta}_j)$ are cluster-robust and OLS variances of β_j, respectively. Similar to $\overline{DE}_y$, one may consider using simple or weighted mean of DE_{β_j}values ($\overline{DE}_\beta$). For example, if we are interested in coefficients of health state descriptors (i.e., β_1 to β_{20}), weights proportional to probability of occurrence of respective severity levels in a sample of randomly allocated blocks can be used for $\overline{DE}_\beta$. According to the valuation protocol, the probability of occurrences of severity levels 2 to 5 in different dimensions range from 0.13 to 0.26 (Table 1).

As an alternative to $V_{Cluster}(\hat{y}_0)$ and $V_{Cluster}(\hat{\beta}_j)$ in equations (6) and (7) respectively, one can use variances estimated for predicted utility value of HS_0, $V_{RE}(\hat{y}_0)$, and β_j, $V_{RE}(\hat{\beta}_j)$, respectively, from the Basic RE model in equation (2). Empirical estimates of design effects from the Singaporean study will be given in Results section.

Equations (6) and (7) assume that participants in the study are selected from the population using simple random sampling. However, a complex survey design such as multistage sampling and/or quota sampling may be used by investigators. In such cases, the design effect for that specific study design should be estimated as per standard survey methodology [16, 17], and then multiply it with the sample size (N) calculated in equation (5).

Sample size estimation approaches for value set studies

Approach 1 - To achieve desired precision for an estimated mean utility value for a particular health state

The primary objective of value set studies is to predict utilities with acceptable precision. Suppose the study team would like to predict utility value of a particular health state HS_0 with tolerated margin of error δ, where δ

is a sufficiently small value (e.g., minimal important difference (MID) for utility values) [18]. That is, the maximum allowable difference between the predicted value and the true value should be less than δ with sufficiently high probability 100×(1-α)%, 0 < α < 1. This is equivalent to targeting the 100×(1-α)% prediction interval (PI) to be ± δ.

Consider the Basic OLS model for estimating the mean utility value of a particular health state, *HS₀*. The 100×(1−α)% PI for mean utility value of *HS₀* is [19]:

$$\left(\hat{y}_0 - Z_{1-\frac{\alpha}{2}}\sqrt{\hat{\sigma}^2\, x_0' C x_0},\ \hat{y}_0 + Z_{1-\frac{\alpha}{2}}\sqrt{\hat{\sigma}^2\, x_0' C x_0}\right), \tag{8}$$

where $\hat{y}_0$ is the predicted utility value for *HS₀*, $x_0' = (1, x_{01}, x_{02}, \ldots, x_{020})$ is the vector representing values of indicator variables for health state descriptor of *HS₀*, $\hat{\sigma}^2$ is the estimate of error variance, and $C = (X'X)^{-1}$.

Thus, for the objective of estimating mean utility value of *HS₀* with desired precision, say 100×(1-α)% PI of $\hat{y}_0$ equal to $\hat{y}_0 \pm \delta$, the sample size N_s can be calculated by solving equation, $\delta = Z_{1-\frac{\alpha}{2}}\sqrt{\hat{\sigma}^2\, x_0' C x_0}$. It can be solved using equation (3) as $\delta = Z_{1-\frac{\alpha}{2}} \times \hat{\sigma} \times \sqrt{10\, x_0' D x_0 / N_s}$ giving:

$$N_s = \frac{10\,\hat{\sigma}^2\, x_0' D x_0\ Z^2_{1-\alpha/2}}{\delta^2}. \tag{9}$$

This approach provides different sample sizes for different health states. The study team may choose the mean of these sample sizes. Using the matrix *D* (Table 2) and x_0 vectors of the directly valued health states in the EQ-VT protocol, it can be shown that $x_0' D x_0$ has mean, standard deviation (SD) and coefficient of variation (CV) of 0.210, 0.062, and 29.3%, respectively. The $x_0' D x_0$ values are same under the dummy and the backward difference coding schemes for any x_0.

Lastly, as equation (9) also assumes independent observations, it needs to be adjusted for the design effect $\overline{DE}_y$ using equation (5), and replacing $x_0' D x_0$ by mean of $x_0' D x_0$ values ($\bar{D}_{x_0}$) of directly values health states, gives

$$N = N_s \times DE = \frac{10\,\hat{\sigma}^2\, \bar{D}_{x_0}\, \overline{DE}_y\ Z^2_{1-\alpha/2}}{\delta^2}. \tag{10}$$

Approach 2 - Empirical approach to achieve desired mean absolute error in prediction of utility values relative to a reference study

This is an empirical approach in which mean absolute error (MAE) defined as [20],

$$MAE_{N_s} = \frac{1}{10N}\sum_{i=1}^{N}\sum_{j=1}^{10}\left|\hat{y}_{ij(N_s)} - y_{ij}\right|, \quad (11)$$

is calculated, where $\hat{y}_{ij(N_s)}$ denote predicted utility values based on a regression model fitted with sample size N_s, and y_{ij} denote observed utility values in a reference study with a larger sample size N, where $N > N_s$. MAE_{N_s} is to be estimated for a large number of replications of random sampling (with replacement) of N_s number of subjects within a reference study with sample size N conducted according to the EQ-VT protocol. A plot of mean MAE_{N_s} over sample size N_s can provide a visual presentation of how the MAE is reduced with an increase in sample size for the model. A sample size N_s^* corresponding to desired MAE, or desired marginal gain in MAE as sample size increases, is selected for the study.

Approach 3 - To assess significance of a regression coefficient of health state descriptors

A value set study is expected to show that individual regression coefficients of health state descriptors (i.e., β_1 to β_{20}) are not different from zero by chance. That is, it requires to test the null hypothesis H_0: $\beta_j = 0$ against the alternative hypothesis H_1: $\beta_j \neq 0$, for all j = 1, 2..., 20, with sufficient statistical power (1-β) and type-I error rate (α).

Considering the Basic OLS model in equation (1), the SE of $\hat{\beta}_j$ is $SE(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 C_{jj}}$, where $\hat{\sigma}^2$ is the estimate of error variance, and C_{jj} is the j^{th} diagonal element of $C = (X'X)^{-1}$ matrix[15]. Using equation (4), $SE(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 D_{jj}/M} = \sqrt{10\hat{\sigma}^2 D_{jj}/N_s}$, where D_{jj} are available from Table 2. Solving the equation, $\beta_j/SE(\hat{\beta}_j) = \beta_j/\sqrt{10\hat{\sigma}^2 D_{jj}/N_s} = (Z_{1-\alpha/2} + Z_{1-\beta})$ will give sample size:

$$N_s = \frac{10\hat{\sigma}^2 D_{jj}(Z_{1-\alpha/2}+Z_{1-\beta})^2}{\beta_j^2}, \quad (12)$$

where $Z_{1-\alpha/2}$ and $Z_{1-\beta}$ are values of standard normal distribution at $1-\alpha/2$ and $1-\beta$, respectively. The $\hat{\sigma}^2$ can be obtained from a reference study (see Results section).

The β_j can be a smallest desired value for the coefficient (e.g., MID for utility values) that the study team would consider meaningful. It is clear that the sample size may be different for different coefficients; one can choose the mean of coefficients ($\bar{\beta}$) for simplicity in the sample size estimation. Substituting $\bar{\beta}$ for β_j and $\bar{D}$ for D_{jj} in equation (12) gives,

$$N_s = \frac{10\,\hat{\sigma}^2\,\bar{D}\,(Z_{1-\alpha/2}+Z_{1-\beta})^2}{\bar{\beta}^2}. \quad (13)$$

Since the sample size formula assumes independent observations, it needs to be adjusted for the design effect $\overline{DE}_\beta$ using equations (5), which gives,

$$N = \frac{10\,\hat{\sigma}^2\,\bar{D}\,\overline{DE}_\beta\,(Z_{1-\alpha/2}+Z_{1-\beta})^2}{\bar{\beta}^2}. \quad (14)$$

Approach 4 - To estimate a regression coefficient of health state descriptors with a desired precision

Alternative to Approach 3, the study team may desire to estimates the regression coefficients with certain precision, say requiring the 100×(1-α)% confidence interval (CI) of β_j to be $\hat{\beta}_j \pm \delta_j$, where δ_j is a sufficiently small number (e.g., MID for utility values). For this objective, the sample size is obtained by solving [19] $\delta_j = Z_{1-\frac{\alpha}{2}} SE(\hat{\beta}_j) = Z_{1-\frac{\alpha}{2}}\sqrt{10\hat{\sigma}^2 D_{jj}/N_s}$. It gives

$$N_s = \frac{10\,\hat{\sigma}^2 D_{jj} Z_{1-\alpha/2}^2}{\delta_j^2}. \quad (15)$$

Alternatively, substituting δ_j by $\bar{\delta}$ = mean of δ_j (j = 1, 2,.., 20), D_{jj} by $\bar{D}$ in equation (15) gives

$$N_s = \frac{10\,\hat{\sigma}^2\,\bar{D} Z_{1-\alpha/2}^2}{\bar{\delta}^2}. \quad (16)$$

Finally, adjusting the sample size for the design effect $\overline{DE}_\beta$ using equation (5), gives

$$N = \frac{10\,\hat{\sigma}^2\,\bar{D}\,\overline{DE}_\beta\, Z_{1-\alpha/2}^2}{\bar{\delta}^2}. \quad (17)$$

Results

Estimates of useful parameters

We used the cTTO utility values of the EQ-5D-5L value set study in the Singapore general population conducted according to the EQ-VT protocol to estimate useful parameters for sample size determination. In a cross-sectional national household survey, members of the general public, each from a different household, were recruited and interviewed by computer-assisted personal interviewing (CAPI). Quotas were set to make the sample representative the general adult Singapore population in distributions of age groups, genders, and ethnic groups. After excluding interviews of poor quality, a total of 1000 participants were used to estimate the EQ-

5D-5L value set for Singapore.

Using the Basic OLS model in equation (1), the estimate of error variance ($\hat{\sigma}^2$) was 0.365. The mean ($\bar{\beta}$) of regression coefficients ($\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_{20}$) of health state descriptor defined using the dummy coding scheme was −0.196. Using the Basic RE model in equation (2), the estimate of between-participant variance ($\hat{\sigma}_\gamma^2$) and the within-participant variance ($\hat{\sigma}_e^2$) were 0.203 and 0.162, respectively. The Basic RE model gave $\bar{\beta}$ −0.204, which was similar to the -0.196 from the Basic OLS model.

Simple and weighted means of design effects $\left(\overline{DE}_\beta\right)$ for $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_{20}$, using equation (7) for cluster-robust variances, were 0.82 and 0.84, respectively, or approximately 0.8. The Basic RE model was more efficient, in the sense that the variance estimates were smaller. Simple and weighted mean design effects based on RE variances were both 0.47, or approximately 0.5. Design effects for the intercept, $\hat{\beta}_0$, were 1.27 and 0.92 based on cluster-robust and RE variances, respectively. Detailed results of the Basic OLS and RE models are presented in Electronic Supplementary Material Table A.2 and A.3 for the dummy and backward difference coding schemes, respectively.

Pearson's correlation between predicted utility values based on the Basic OLS and RE models was 0.999 in the Singapore study; and mean absolute error in prediction using both the models was 0.499. Mean design effects $\left(\overline{DE}_{y_0}\right)$ for predicted values using equation (6) were approximately 1.15 and 0.75 based on cluster-robust and RE variances respectively. Weighted means of design effects were similar to their unweighted mean design effects. Mean utility values of individual health states in the Singaporean value set study along with their predicted values and their standard errors (square root of variances) using the Basic OLS and RE models are presented in Electronic Supplementary Material Table A.4.

Numerical illustration

For illustration, an MID of 0.05 as reported by Nolan et al. [21] is considered for utility values and regression coefficients of health state descriptors.

Approach 1 - To achieve desired precision of an estimated mean utility value for a particular health state

Suppose the study team is interested in calculating a sample size such that the predicted utility value of '12334' health state, which has $x_0' D x_0$ value approximately equal to the mean of $x_0' D x_0$ values (i.e., $\overline{D}_{x_0} = 0.210$) of all directly valued health states in the EQ-VT protocol, is not far from its true value by more than a MID for utility

values. That is, the half-width of the 95% PI of the estimated utility value equal to δ = 0.05 (MID). The sample size can be calculated using equation (10), assuming $\hat{\sigma}^2 = 0.365$ and design effect $\overline{DE}_y$ = 0.75 (based on the Basic RE model), as $N = 10\,\hat{\sigma}^2\,\overline{D}_{x_0}\overline{DE}_y\,Z^2_{1-\alpha/2}/\delta^2$ = 10×0.365×0.210×0.75×(1.960)^2/(0.05)^2 = 2.2084/0.0025 ≈ 883 participants.

The solid line in Fig. 1 shows the sample sizes calculated as illustrated above for Approach 1 for different MID (y-axis on left). It shows a rapid decline in the curve with increase in sample size from 100 to 400 participants. However, it decreases by a trivial magnitude with further increase of sample size after 400.

Approach 2 - Empirical approach to achieve desired mean absolute error in prediction of utility values relative to a reference study

For this approach, Fig. 1 shows the mean of MAE_{N_s} (y-axis on right) for 100 random samples (with replacement) of sample sizes 100 to 900 participants based on the Basic OLS model using the Singapore value set study data (N=1000). The MAE decreased rapidly between the sample sizes of 100 to 300. After that, reduction in the slope is trivial beyond the sample size of 300. That is, gain in prediction accuracy was trivial when the sample size was more than 300 participants. Thus, one may consider 300 participants as a sample size for developing a value set.

Approach 3 - To assess significance of a regression coefficient of health state descriptors

The sample size required to test H_0: $\bar{\beta} = 0$ against H_1: $\bar{\beta} \neq 0$, assuming the true value of $\bar{\beta} = 0.05$ (MID) and $\hat{\sigma}^2$ = 0.365, at 5% two-sided level of significance (α) and 80% statistical power (1−β), can be calculated using equation (14), by inserting value of $(Z_{1-\alpha/2} + Z_{1-\beta})^2$ = (1.960 + 0.842)^2 = 7.851, $\overline{D} = 0.131$, and adjusting for design effect $\overline{DE}_\beta = 0.5$ (based on the Basic RE model), as $N = 10\,\hat{\sigma}^2\,\overline{D}\,\overline{DE}_\beta\,(Z_{1-\alpha/2} + Z_{1-\beta})^2/\bar{\beta}^2$ = 10×0.365×0.131×0.5×7.851/(0.05)^2 = 1.8770/0.0025 ≈ 751 participants. Fig. 1 shows sample sizes corresponding to values of $\bar{\beta}$ (y-axis on left) at 5% of α and 80% statistical power.

Approach 4 - To estimate a regression coefficient of health state descriptors with a desired precision

To estimate a regression coefficient with its 95% CI of half-width equal to $\bar{\delta} = 0.05$ (MID), using equation (17) with $\hat{\sigma}^2$ = 0.365, $\overline{D} = 0.131$ and $\overline{DE}_\beta = 0.5$ (as in previous illustration), the required sample size is $N =$

$10\ \hat{\sigma}^2\ \overline{D}\ \overline{DE}_\beta\ Z^2_{1-\alpha/2}/\bar{\delta}^2$ = 10×0.365×0.131×0.5×(1.960)²/(0.05)² = 0.9184/0.0025 ≈ 367 participants. Fig. 1 shows sample sizes corresponding to values of $\bar{\delta}$ (y-axis on left) for 95% CI of a regression coefficient.

The backward difference coding scheme used for health state descriptors has no impact on sample size estimation in Approach 1 and 2, and has only trivial impact in Approach 3 and 4 (Results not shown).

Discussion

We illustrated four approaches for estimating sample size for health states valuation studies using the EQ-VT protocol – two based on prediction error for utility values, and two based on regression coefficients of an OLS regression model for utility values. Among the first two approaches which are directly related to the objective of value set studies to predict utility values with acceptable precision, Approach 1 based on PIs of predicted utility values has a few advantages over Approach 2 based on MAEs. Using Approach 1, one can control the prediction error with desired level of confidence (probability), whereas Approach 2 is only based on the mean prediction error. Furthermore, Approach 2 requires access to data of another value set study with sufficiently large sample size, and the size of a new study is chosen relative to this bigger study. The last two approaches are also relevant for as they are directly related to the analysis strategy and help to achieve an overall objective of developing a value set with acceptable precision indirectly.

A few other sample size estimation approaches, based on comparing mean utility values of two different health states and on estimating utility values with desired precision, have been used in EQ-5D-3L value set studies [22, 23]. However, these approaches were not directly linked to the analysis strategy, which was to develop a regression model for estimating utility values, in value set studies. The approaches did not involve the model used to predict utility values; therefore, cannot guarantee the precision of utility values estimated using the model. The study team should choose an approach which is most relevant to the study objective, analysis strategy, and outcome.

The key objective of the article is to introduce justified approaches for calculating sample size for value set studies. We are neither recommending any modeling approach, nor presenting a value set. Thus, no serious attempt has been made to improve the model fitting, other than incorporating an intercept and 20 indictor variables for health states descriptors in the models. These basic models have not included any interaction terms, which might be useful for validity of the model assumptions and performance. This could be also a possible reason for a high value of MAE using Approach 2, even at sample size of 1000 participants. How to improve the

analysis models and reduce the prediction error in utility values is a complex issue beyond the scope of the present manuscript, potentially related to valuation methodology as well as modeling technique. This may require research. Furthermore, in the illustrations of sample size determination approaches, we used 0.05 as the MID for utility values only for the purpose of illustrating the numerical applications. In some realistic situations such as those in the illustration for Approaches 1 and 3, a sample size of 1000 participants was about right. However, they were based on specific assumptions and specific level of precision, which may vary from study to study. Users should estimate the sample size according to their own requirements.

We explored how commonly used approaches for statistical inference in clustered data – the cluster-robust estimator and RE model – affect the sample size estimation. Both approaches showed deflation in variances for regression coefficients of health states descriptors when accounted for within-participant clustering effect. It is not surprising to see such results when health states in each block are spread over from very mild to extreme severity; that is, the within-person correlation among health state assignments is negative [13, 14]. Regarding the choice between clustered-robust estimator and RE model approaches for estimating the design effect, it is not so straightforward, and depends on data and study team's preference for the model selection. The cluster-robust estimator approach requires fewer assumptions compared to the RE model approach [24], but it has a larger design effect. The RE model approach is more efficient and leads to smaller SEs and design effect compared to the other approach, but there is no guarantee that the model assumptions hold. Nevertheless, as reported in the Results section, the Pearson's correlation between predicted utility values based on the Basic OLS and RE models was close to 1 in the Singapore study, meaning both the models gave practically the same results.

It is not uncommon to exclude some participants' data in value set studies if they seem logically incorrect. Currently, there is no consensus on who should be excluded from the analysis [25]. However, value set studies usually excludes participants who seem to have not understood the valuation task, and thus provided illogical values. For example, participants who have given the same value to all health states or given only negative values to all health states can be excluded. In the EQ-5D-3L value set studies conducted in various countries, the proportion of unusable data ranged from 0-50% depending on the exclusion criteria used [26]. Thus, depending on estimated proportion of unusable data, say p, the sample size should be increased in that proportion, by multiplying it to $1/(1-p)$ to account for unusable data. Similar adjustments should be made in the sample size to account for incomplete interviews, exclusion of data due to interviewer learning effect, and other possible reasons which could potentially lead to data exclusion [27].

Conclusions

We proposed sample size determination methods for estimating utility values of health states with acceptable precision. Capitalizing on the EQ-5D-5L valuation protocol, we calculated the inverse of the cross-product matrix, which is critical for the sample size estimation. Based on a Singaporean study, we estimated several parameters that are useful for the sample size estimation. We proposed four sample size estimation approaches which can help to decide an appropriate sample size for a value set study. These approaches will be useful for developing value sets with a scientifically justified sample size, and thus improve their quality.

Compliance with Ethical Standards

Disclosure of potential conflict of interest

Funding: Financial support for this study was provided entirely by a Health Services Research Competitive Research Grant (HSRG/0038/2013) from the National Medical Research Council, Singapore. The last author (YBC) was supported by the National Research Foundation, Singapore, under its Clinician Scientist Award (NMRC/CSA/0039/2012) administered by the Singapore Ministry of Health's National Medical Research Council. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Conflict of interest: The authors declare that they have no conflict of interest.

Research involving Human Participants and/or Animals

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

Informed consent

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

References

[1] Longworth, L., Yang, Y., Young, T., Mulhern, B., Alava, M.H., Mukuria, C., et al (2014). Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. Health Technology Assessment, 18(9), 1-224. doi:10.3310/hta18090.

[2] Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., et al. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Quality of life research, 20(10), 1727-1736. doi: 10.1007/s11136-011-9903-x.

[3] Oppe, M., Rand-Hendriksen, K., Shah, K., Ramos-Goñi, J.M., Luo, N. (2016). EuroQol Protocols for Time Trade-Off Valuation of Health Outcomes. PharmacoEconomics, 1-12. doi: 10.1007/s40273-016-0404-1.

[4] Oppe, M., Devlin, N.J., van Hout, B., Krabbe, P.F., de Charro, F. (2014). A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value in Health, 17(4), 445-453. doi: 10.1016/j.jval.2014.04.002.

[5] EuroQol Group: EQ-5D-5L Value Sets. (2016). http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-5l-value-sets.html. Accessed 17 March 2017.

[6] Devlin, N.J., Krabbe, P.F. (2013). The development of new research methods for the valuation of EQ-5D-5L. The European Journal of Health Economics, 14(Suppl 1), 1. doi: 10.1007/s10198-013-0502-3.

[7] EuroQol Group: EQ-5D-5L User Guide. (2015). http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/EQ-5D-5L_UserGuide_2015.pdf. Accessed 06 Febraury 2016.

[8] Luo, N., Wang, P., Thumboo, J., Lim, Y.-W., Vrijhoef, H.J. (2014). Valuation of EQ-5D-3L health states in Singapore: modeling of time trade-off values for 80 empirically observed health states. Pharmacoeconomics, 32(5), 495-507. doi: 10.1007/s40273-014-0142-1.

[9] Williams, R.L. (2000). A note on robust variance estimation for cluster-correlated data. Biometrics, 56(2), 645-646. doi: 10.1111/j.0006-341X.2000.00645.x.

[10] Verbeke, G., Molenberghs, G. (2009). Linear mixed models for longitudinal data. New York: Springer Science & Business Media.

[11] Oppe, M., Devlin, N.J., Szende, A. (2007). EQ-5D value sets: inventory, comparative review and user guide. The Netherlands: Springer.

[12] Chen, X., Ender, P., Michaelis, M., Wells, C. (2003). Regression with Stata. http://www.ats.ucla.edu/stat/stata/webbooks/reg/default.htm. Accessed 17 March 2017.

[13] Parzen, M., Lipsitz, S.R., Dear, K.B. (1998). Does clustering affect the usual test statistics of no treatment effect in a randomized clinical trial? Biometrical journal, 40(4), 385-402. doi: 10.1002/(SICI)1521-4036(199808)40:4.

[14] Kahan, B.C., Morris, T.P. (2013). Assessing potential sources of clustering in individually randomised trials. BMC medical research methodology, 13(1), 1. doi: 10.1186/1471-2288-13-58.

[15] Donner, A., Klar, N. (2000). Design and analysis of cluster randomization trials in health research. London: Wiley.

[16] Heeringa, S. G., West, B. T., & Berglund, P. A. (2010). Applied survey data analysis: CRC Press.

[17] Lee, E. S., & Forthofer, R. N. (2005). Analyzing complex survey data (vol. 71): Sage Publications.

[18] Coretti, S., Ruggeri, M., McNamee, P. (2014). The minimum clinically important difference for EQ-5D index: a critical review. Expert review of pharmacoeconomics & outcomes research, 14(2), 221-33. doi: 10.1586/14737167.2014.894462.

[19] Montgomery, D.C., Peck, E.A., Vining, G.G. (2015). Introduction to linear regression analysis. New York: John Wiley & Sons.

[20] Lamers, L.M., McDonnell, J., Stalmeier, P.F., Krabbe, P.F., Busschbach, J.J. (2006). The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health economics, 15(10), 1121-1132. doi: 10.1002/hec.1124.

[21] Nolan, C.M., Longworth, L., Lord, J., Canavan, J.L., Jones, S.E., Kon, S.S., et al. (2016). The EQ-5D-5L health status questionnaire in COPD: validity, responsiveness and minimum important difference. Thorax, 71(6), 493-500. doi: 10.1136/thoraxjnl-2015-207782.

[22] Dolan, P., Dudex, C., Kind, P., Williams, A. (1994). The measurement and valuation of health, The first report on the main survey. https://www.york.ac.uk/media/che/documents/reports/MVH%20First%20Report.pdf (1994). Accessed 17 March 2017.

[23] Chevalier, J., de Pouvourville, G. (2013). Valuing EQ-5D using time trade-off in France. The European Journal of Health Economics, 14(1), 57-66. doi: 10.1007/s10198-011-0351-x.

[24] Cameron, A.C., Miller, D.L. (2015). A practitioner's guide to cluster-robust inference. Journal of Human Resources, 50(2), 317-372. doi: 10.3368/jhr.50.2.317.

[25] Lamers, L.M., Stalmeier, P.F., Krabbe, P.F., Busschbach, J.J. (2006). Inconsistencies in TTO and VAS values for EQ-5D health states. Medical decision making, 26(2), 173-181. doi: 10.1177/0272989X06286480.

[26] Devlin, N.J., Hansen, P., Kind, P., Williams, A. (2003). Logical inconsistencies in survey respondents' health state valuations-a methodological challenge for estimating social tariffs. Health economics, 12(7), 529-544. doi: 10.1002/hec.741.

[27] Purba, F.D., Hunfeld, J.A., Iskandarsyah, A., Fitriana, T.S., Sadarjoen, S.S., Passchier, J., et al. (2016). Employing quality control and feedback to the EQ-5D-5L valuation protocol to improve the quality of data collection. Quality of Life Research, 1-12. doi: 10.1007/s11136-016-1445-9.

Figure captions

Fig. 1 Comparison of sample sizes estimated using different approaches.

PI: Prediction interval. CI: Confidence interval. Refer to y-axis on left for Approach 1, 3 and 4, and y-axis on right for Approach 3. For Approach 2, mean absolute errors are based on 1000 replications for sample sizes from 100 to 900 participants (each of the participant valued 10 health states), whereas they are based on a single sample for the sample size of 1000 participants.

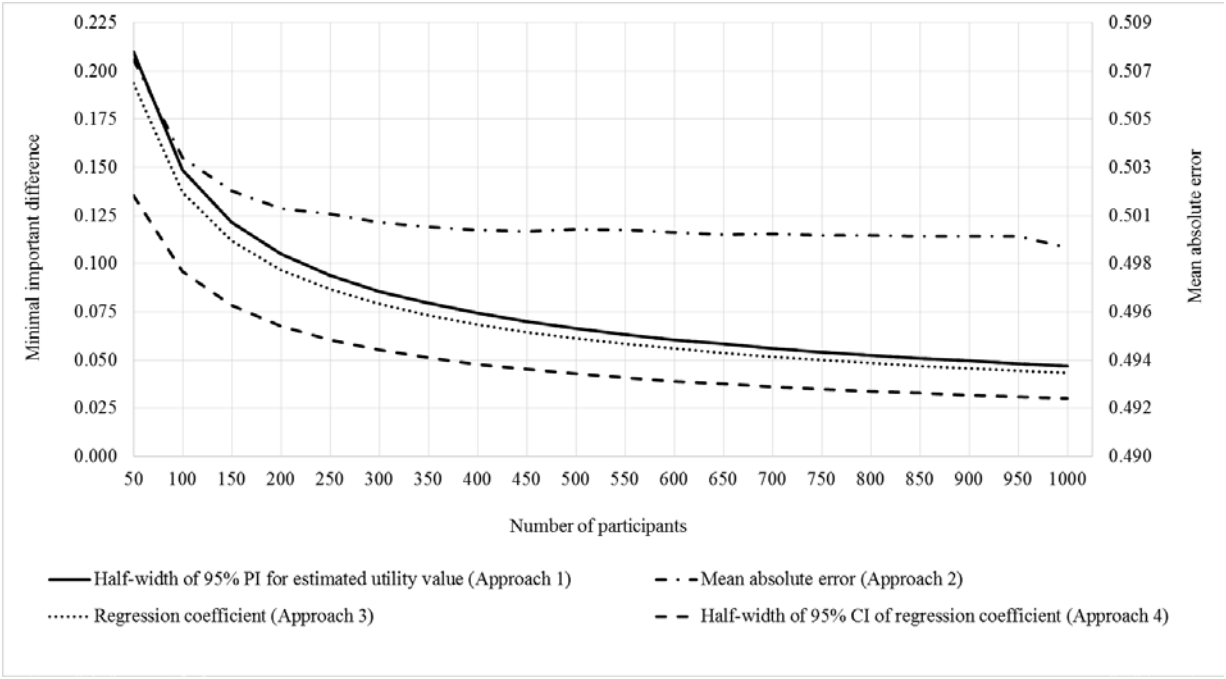

0.225
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
Minimal important difference
0.509
0.507
0.505
0.503
0.501
0.498
0.496
0.494
0.492
0.490
Mean absolute error
50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850 900 950 1000
Number of participants
Half-width of 95% PI for estimated utility value (Approach 1)
Mean absolute error (Approach 2)
Regression coefficient (Approach 3)
Half-width of 95% CI of regression coefficient (Approach 4)

Table 1 Unique blocks of EQ-5D-5L health states for the direct valuation with composite time trade-off method.

	Health state						Health state				
Block	MO	SC	UA	PD	AD	Block	MO	SC	UA	PD	AD
1	1	1	2	2	1	6	1	2	1	1	2
	1	1	2	3	5		1	1	2	1	2
	5	4	2	3	1		4	4	5	5	3
	5	1	4	5	1		2	1	3	4	5
	3	4	5	1	5		3	4	2	4	4
	3	5	2	4	5		2	3	1	5	2
	1	2	5	1	4		4	3	5	1	4
	4	5	1	4	4		5	5	4	2	4
	1	2	1	1	1		2	1	1	1	1
	5	5	5	5	5		5	5	5	5	5
2	1	2	5	4	3	7	1	3	1	2	2
	1	2	1	2	1		2	4	5	5	3
	4	3	5	4	2		5	1	1	5	2
	3	4	1	5	5		1	1	4	2	5
	5	2	2	1	5		2	2	4	3	4
	4	5	1	3	3		4	2	1	1	5
	3	2	4	4	3		3	5	3	3	2
	2	3	5	1	4		4	5	4	1	3
	1	1	2	1	1		1	1	2	1	1
	5	5	5	5	5		5	5	5	5	5
3	4	5	2	3	3	8	3	3	2	5	3
	5	5	2	3	3		2	3	2	4	2
	3	1	5	2	5		2	4	3	4	2
	5	2	4	5	5		3	2	3	1	4
	1	2	2	4	4		1	2	3	3	4
	1	3	3	1	3		2	1	3	3	4
	2	5	1	2	2		5	5	2	2	5
	1	1	4	2	1		5	3	4	1	2
	2	1	1	1	1		1	1	1	1	2
	5	5	5	5	5		5	5	5	5	5
4	2	1	1	1	2	9	1	1	4	1	4
	1	4	5	5	4		2	5	3	3	1
	1	2	5	1	3		2	5	2	2	2
	4	4	3	4	5		2	1	4	4	4
	1	2	3	4	4		3	1	5	1	4
	5	3	2	2	1		5	3	2	4	3
	5	4	3	4	2		5	3	2	4	4
	4	4	1	2	5		3	5	1	4	3
	1	1	1	2	1		1	1	1	2	1
	5	5	5	5	5		5	5	5	5	5
5	4	3	3	1	5	10	1	1	1	2	2
	5	4	1	5	3		5	2	3	3	5
	5	2	4	3	1		3	5	3	1	1
	2	4	4	4	3		4	3	5	5	5
	1	4	1	1	3		2	4	4	4	5
	3	1	5	2	4		1	3	2	2	4
	1	5	1	5	1		3	4	2	3	2
	2	1	3	1	5		4	2	3	2	1
	1	1	1	1	2		1	2	1	1	1
	5	5	5	5	5		5	5	5	5	5

EQ-5D-5L: 5-Level EuroQol 5-Dimension; MO: Mobility, SC: Self-care; UA: Usual activities; PD: Pain/discomfort; AD: Anxiety/depression.

Table 2 *D* matrix based on indicator variables defined using the dummy coding scheme for EQ-5D-5L health state descriptors.

	INT	MO2	MO3	MO4	MO5	SC2	SC3	SC4	SC5	UA2	UA3	UA4	UA5	PD2	PD3	PD4	PD5	AD2	AD3	AD4	AD5
INT	0.131																				
MO2	-0.037	0.117																			
MO3	-0.022	0.052	0.129																		
MO4	-0.029	0.057	0.058	0.155																	
MO5	-0.013	0.056	0.059	0.066	0.130																
SC2	-0.051	0.022	0.007	0.000	-0.008	0.111															
SC3	-0.011	-0.016	-0.011	-0.052	-0.033	0.045	0.148														
SC4	-0.020	-0.012	-0.024	-0.034	-0.021	0.051	0.067	0.140													
SC5	-0.019	-0.013	-0.025	-0.038	-0.042	0.052	0.068	0.074	0.124												
UA2	-0.050	0.011	-0.008	0.023	-0.018	0.015	-0.031	-0.001	0.001	0.125											
UA3	-0.041	-0.012	-0.014	-0.001	-0.013	-0.005	-0.018	-0.004	-0.001	0.066	0.136										
UA4	-0.039	-0.018	-0.011	-0.002	-0.036	0.006	0.005	0.013	0.019	0.063	0.061	0.141									
UA5	-0.028	-0.001	-0.010	0.001	-0.012	0.002	-0.014	0.000	-0.006	0.053	0.052	0.056	0.123								
PD2	-0.051	0.011	0.006	0.006	0.004	0.010	-0.004	-0.006	-0.012	-0.002	0.010	-0.006	0.007	0.101							
PD3	-0.007	-0.020	-0.008	-0.020	-0.022	-0.015	0.017	-0.015	-0.023	-0.037	-0.035	-0.017	0.004	0.036	0.150						
PD4	0.000	-0.019	-0.008	-0.008	-0.010	-0.013	-0.007	-0.029	-0.018	-0.024	-0.018	-0.018	0.005	0.034	0.059	0.114					
PD5	-0.035	-0.002	0.003	0.012	-0.034	0.006	-0.016	-0.029	-0.025	0.022	0.021	0.009	-0.019	0.041	0.037	0.040	0.138				
AD2	-0.063	-0.008	-0.006	0.007	-0.001	0.007	-0.028	-0.012	-0.009	0.019	0.015	0.016	0.005	0.009	-0.006	-0.015	0.008	0.126			
AD3	-0.050	0.006	-0.008	-0.010	0.013	-0.013	-0.031	-0.033	-0.028	-0.001	0.011	-0.015	-0.019	0.022	-0.011	-0.021	-0.001	0.068	0.156		
AD4	-0.045	0.000	-0.010	0.000	0.010	-0.010	-0.020	-0.007	-0.006	-0.007	-0.009	-0.022	-0.040	0.009	-0.011	-0.024	0.007	0.063	0.079	0.135	
AD5	-0.037	-0.009	-0.018	-0.027	-0.020	-0.003	0.003	-0.008	-0.009	-0.010	-0.016	-0.012	-0.034	0.006	0.007	-0.013	-0.012	0.059	0.077	0.071	0.124

EQ-5D-5L: 5-Level EuroQol 5-Dimension; INT: Intercept; MO2 to MO4, SC2 to SC5, UA2 to UA5, PD2 to PD5, and AD2 to AD5 represent elements corresponding to indicator variables for Self-care, Usual activity, Pain/discomfort, and Anxiety/depression dimension at severity level 2 to 4 (Reference: severity level 1), respectively.

Electronic Supplementary Material

Article title: Sample size for EQ-5D-5L value set studies

Journal: Quality of Life Research

Author names: Mihir Gandhi, Ying Xu, Nan Luo, Yin Bun Cheung

Corresponding author: Mihir Gandhi, Head of Biostatistics, Singapore Clinical Research Institute; email: mihir.gandhi@scri.edu.sg

Table A.1 *D* matrix based on indicator variables defined using the backward difference coding scheme for EQ-5D-5L health state descriptors.

	INT	MO2	MO3	MO4	MO5	SC2	SC3	SC4	SC5	UA2	UA3	UA4	UA5	PD2	PD3	PD4	PD5	AD2	AD3	AD4	AD5
INT	0.131																				
MO2	-0.037	0.117																			
MO3	0.015	-0.065	0.141																		
MO4	-0.007	0.005	-0.075	0.167																	
MO5	0.016	-0.001	0.002	-0.090	0.154																
SC2	-0.051	0.022	-0.015	-0.007	-0.008	0.111															
SC3	0.040	-0.038	0.020	-0.034	0.027	-0.066	0.169														
SC4	-0.010	0.004	-0.017	0.032	-0.007	0.006	-0.087	0.154													
SC5	0.001	-0.001	0.001	-0.004	-0.017	0.001	0.000	-0.066	0.115												
UA2	-0.050	0.011	-0.019	0.031	-0.042	0.015	-0.046	0.030	0.002	0.125											
UA3	0.009	-0.023	0.017	-0.017	0.029	-0.020	0.033	-0.016	0.002	-0.060	0.130										
UA4	0.002	-0.006	0.009	-0.004	-0.021	0.011	0.012	-0.006	0.003	-0.003	-0.072	0.155									
UA5	0.011	0.018	-0.017	0.002	0.021	-0.003	-0.015	0.006	-0.012	-0.010	0.000	-0.075	0.151								
PD2	-0.051	0.011	-0.004	-0.001	-0.001	0.010	-0.015	-0.001	-0.007	-0.002	0.012	-0.016	0.013	0.101							
PD3	0.045	-0.030	0.016	-0.011	0.000	-0.025	0.047	-0.031	-0.001	-0.035	-0.011	0.035	0.007	-0.065	0.179						
PD4	0.007	0.001	-0.001	0.012	0.000	0.002	-0.026	0.010	0.019	0.012	0.004	-0.018	0.002	-0.002	-0.089	0.146					
PD5	-0.035	0.017	-0.006	0.010	-0.045	0.019	-0.027	0.009	-0.007	0.046	-0.007	-0.013	-0.050	0.007	-0.029	-0.052	0.172				
AD2	-0.063	-0.008	0.002	0.013	-0.008	0.007	-0.035	0.016	0.003	0.019	-0.004	0.002	-0.011	0.009	-0.015	-0.009	0.023	0.126			
AD3	0.013	0.014	-0.016	-0.015	0.031	-0.020	0.017	-0.019	0.001	-0.020	0.016	-0.028	0.008	0.013	-0.018	-0.001	-0.003	-0.058	0.147		
AD4	0.005	-0.006	0.003	0.013	-0.012	0.003	0.007	0.016	-0.004	-0.006	-0.014	0.014	-0.015	-0.014	0.014	-0.003	0.010	-0.005	-0.073	0.135	
AD5	0.008	-0.009	0.001	-0.019	-0.003	0.006	0.016	-0.024	-0.002	-0.003	-0.004	0.018	-0.004	-0.003	0.021	-0.007	-0.029	-0.004	0.003	-0.063	0.117

EQ-5D-5L: 5-Level EuroQol 5-Dimension; INT: Intercept; MO2 to MO4, SC2 to SC5, UA2 to UA5, PD2 to PD5, and AD2 to AD5 represent elements corresponding to indicator variables for Self-care, Usual activity, Pain/discomfort, and Anxiety/depression dimension at severity level 2 to 4 (Reference: previous severity level), respectively.

Table A.2 Summary of Basic OLS and RE models for estimating utility values, using health state descriptors defined with the dummy coding scheme, based on the EQ-5D-5L value set study in the Singapore.

Health state descriptor	Basic OLS Model Coefficient (SE; $SE_{Cluster}$)	Basic RE Model Coefficient (SE_{RE})
Mobility Level 2	-0.072 (0.0207; 0.0193)	-0.093 (0.0141)
Mobility Level 3	-0.127 (0.0217; 0.0189)	-0.141 (0.0147)
Mobility Level 4	-0.253 (0.0238; 0.0198)	-0.269 (0.0161)
Mobility Level 5	-0.294 (0.0217; 0.0169)	-0.298 (0.0147)
Self-care Level 2	-0.118 (0.0202; 0.0193)	-0.131 (0.0142)
Self-care Level 3	-0.152 (0.0233; 0.0198)	-0.160 (0.0159)
Self-care Level 4	-0.288 (0.0226; 0.0246)	-0.269 (0.0158)
Self-care Level 5	-0.269 (0.0213; 0.0151)	-0.271 (0.0143)
Usual activities Level 2	-0.081 (0.0214; 0.0195)	-0.090 (0.0147)
Usual activities Level 3	-0.150 (0.0223; 0.0259)	-0.156 (0.0157)
Usual activities Level 4	-0.238 (0.0227; 0.0176)	-0.240 (0.0155)
Usual activities Level 5	-0.182 (0.0211; 0.0156)	-0.190 (0.0143)
Pain/discomfort Level 2	-0.072 (0.0192; 0.0173)	-0.086 (0.0132)
Pain/discomfort Level 3	-0.145 (0.0234; 0.0189)	-0.140 (0.0159)
Pain/discomfort Level 4	-0.260 (0.0203; 0.0202)	-0.269 (0.0142)
Pain/discomfort Level 5	-0.332 (0.0226; 0.0188)	-0.343 (0.0153)
Anxiety/depression Level 2	-0.084 (0.0215; 0.0230)	-0.105 (0.0154)
Anxiety/depression Level 3	-0.174 (0.0239; 0.0273)	-0.185 (0.0171)
Anxiety/depression Level 4	-0.312 (0.0222; 0.0222)	-0.325 (0.0156)
Anxiety/depression Level 5	-0.324 (0.0213; 0.0177)	-0.322 (0.0144)
Intercept	0.840 (0.0219; 0.0247)	0.868 (0.0210)
R^2 (Ordinary least square regression) = 28.9%		
Mean error ($\hat{\sigma}^2$) = Between-participant error ($\hat{\sigma}_u^2$) + Within-participant error ($\hat{\sigma}_e^2$) = 0.2032 + 0.1616 = 0.3653		

EQ-5D-5L: 5-Level EuroQol 5-Dimension; OLS: Ordinary least square; RE: Random effects; SE: Ordinary least square standard error; $SE_{Cluster}$: Cluster-robust standard error; SE_{RE}: Random effect standard error.

Table A.3 Summary of Basic OLS and RE models for estimating utility values, using health state descriptors defined with the backward difference coding scheme, based on the EQ-5D-5L value set study in the Singapore.

Health state descriptor	Basic OLS Model Coefficient (SE; $SE_{Cluster}$)	Basic RE Model Coefficient (SE_{RE})
Mobility Level 2	-0.072 (0.0207; 0.0193)	-0.093 (0.0141)
Mobility Level 3	-0.055 (0.0227; 0.0177)	-0.048 (0.0154)
Mobility Level 4	-0.126 (0.0248; 0.0222)	-0.128 (0.0170)
Mobility Level 5	-0.041 (0.0237; 0.0221)	-0.029 (0.0164)
Self-care Level 2	-0.118 (0.0202; 0.0193)	-0.131 (0.0142)
Self-care Level 3	-0.034 (0.0249; 0.0211)	-0.029 (0.0169)
Self-care Level 4	-0.136 (0.0237; 0.0258)	-0.109 (0.0166)
Self-care Level 5	0.019 (0.0204; 0.0243)	-0.001 (0.0146)
Usual activities Level 2	-0.081 (0.0214; 0.0195)	-0.090 (0.0147)
Usual activities Level 3	-0.069 (0.0219; 0.0239)	-0.065 (0.0156)
Usual activities Level 4	-0.088 (0.0237; 0.0262)	-0.084 (0.0168)
Usual activities Level 5	0.056 (0.0235; 0.0217)	0.050 (0.0164)
Pain/discomfort Level 2	-0.072 (0.0192; 0.0173)	-0.086 (0.0132)
Pain/discomfort Level 3	-0.073 (0.0256; 0.0242)	-0.055 (0.0176)
Pain/discomfort Level 4	-0.115 (0.0231; 0.0238)	-0.129 (0.0161)
Pain/discomfort Level 5	-0.073 (0.0251; 0.0291)	-0.073 (0.0180)
Anxiety/depression Level 2	-0.084 (0.0215; 0.0230)	-0.105 (0.0154)
Anxiety/depression Level 3	-0.091 (0.0231; 0.0239)	-0.080 (0.0163)
Anxiety/depression Level 4	-0.138 (0.0222; 0.0236)	-0.140 (0.0154)
Anxiety/depression Level 5	-0.012 (0.0208; 0.0219)	0.002 (0.0146)
Intercept	0.840 (0.0219; 0.0247)	0.868 (0.0210)
R^2 (Ordinary least square regression) = 28.9%		
Mean error ($\hat{\sigma}^2$) = Between-participant error ($\hat{\sigma}_u^2$) + Within-participant error ($\hat{\sigma}_e^2$) = 0.2032 + 0.1616 = 0.3653		

EQ-5D-5L: 5-Level EuroQol 5-Dimension; OLS: Ordinary least square; RE: Random effects; SE: Ordinary least square standard error; $SE_{Cluster}$: Cluster-robust standard error; SE_{RE}: Random effect standard error.

Table A.4 Summary of observed and predicted utility values using the Basic OLS and RE models of EQ-5D-5L health states elicited using composite time trade-off method in the Singapore value set study.

Health States	n	Observed Mean (SD)	Basic OLS Model Predicted Value (SE; $SE_{Cluster}$)	Basic RE Model Predicted Value (SE_{RE})
55555	1000	-0.516 (0.566)	-0.562 (0.017; 0.020)	-0.556 (0.018)
11121	214	0.893 (0.247)	0.767 (0.022; 0.020)	0.782 (0.021)
11211	204	0.859 (0.280)	0.759 (0.024; 0.023)	0.778 (0.022)
11112	201	0.844 (0.318)	0.756 (0.022; 0.018)	0.763 (0.021)
12111	195	0.770 (0.417)	0.721 (0.023; 0.022)	0.737 (0.021)
21111	186	0.848 (0.346)	0.768 (0.025; 0.025)	0.775 (0.023)
21112	110	0.765 (0.392)	0.684 (0.024; 0.025)	0.670 (0.022)
12513	110	0.321 (0.632)	0.365 (0.027; 0.028)	0.362 (0.024)
53221	110	0.160 (0.672)	0.241 (0.028; 0.029)	0.234 (0.024)
12344	110	-0.090 (0.669)	-0.001 (0.027; 0.031)	-0.013 (0.023)
44125	110	-0.150 (0.667)	-0.097 (0.030; 0.036)	-0.078 (0.025)
54342	110	-0.281 (0.656)	-0.236 (0.032; 0.041)	-0.230 (0.027)
14554	110	-0.290 (0.644)	-0.276 (0.029; 0.034)	-0.259 (0.025)
44345	110	-0.333 (0.610)	-0.435 (0.028; 0.039)	-0.418 (0.025)
11414	104	0.310 (0.591)	0.289 (0.027; 0.028)	0.303 (0.023)
25222	104	0.263 (0.667)	0.262 (0.033; 0.038)	0.224 (0.027)
25331	104	0.218 (0.654)	0.203 (0.030; 0.029)	0.209 (0.025)
31514	104	0.166 (0.608)	0.218 (0.027; 0.029)	0.212 (0.024)
35143	104	-0.008 (0.643)	-0.043 (0.027; 0.034)	-0.059 (0.024)
53243	104	-0.043 (0.660)	0.009 (0.031; 0.038)	0.001 (0.026)
21444	104	-0.059 (0.657)	-0.121 (0.029; 0.036)	-0.135 (0.025)
53244	104	-0.100 (0.648)	-0.259 (0.028; 0.039)	-0.275 (0.025)
21315	103	0.211 (0.634)	0.377 (0.028; 0.031)	0.413 (0.024)
14113	103	0.151 (0.704)	0.294 (0.027; 0.028)	0.297 (0.024)
52431	103	-0.011 (0.643)	0.238 (0.029; 0.024)	0.255 (0.024)
31524	103	-0.031 (0.677)	0.146 (0.030; 0.027)	0.126 (0.025)
43315	103	-0.077 (0.654)	0.045 (0.031; 0.031)	0.059 (0.026)
15151	103	-0.114 (0.670)	-0.039 (0.029; 0.031)	-0.039 (0.025)
24443	103	-0.187 (0.654)	-0.193 (0.028; 0.030)	-0.188 (0.024)
54153	103	-0.226 (0.640)	-0.249 (0.030; 0.035)	-0.228 (0.025)
12121	102	0.746 (0.369)	0.649 (0.024; 0.023)	0.651 (0.022)
13122	102	0.619 (0.455)	0.532 (0.028; 0.025)	0.517 (0.024)
35332	102	0.152 (0.606)	0.205 (0.028; 0.027)	0.220 (0.024)
23514	102	0.106 (0.618)	0.145 (0.029; 0.030)	0.145 (0.024)
22434	102	0.093 (0.644)	0.130 (0.030; 0.027)	0.122 (0.025)
11425	102	0.087 (0.645)	0.105 (0.031; 0.035)	0.093 (0.026)
45133	102	0.045 (0.634)	-0.046 (0.031; 0.036)	-0.060 (0.026)
12543	102	0.038 (0.677)	0.121 (0.029; 0.031)	0.101 (0.024)
42115	102	0.036 (0.637)	0.023 (0.029; 0.026)	0.026 (0.024)
32443	102	-0.008 (0.628)	-0.078 (0.031; 0.037)	-0.098 (0.026)
52215	102	-0.022 (0.641)	0.065 (0.031; 0.034)	0.055 (0.026)
43542	102	-0.097 (0.629)	-0.002 (0.030; 0.033)	0.003 (0.025)
45413	102	-0.100 (0.610)	-0.095 (0.032; 0.034)	-0.096 (0.027)
34155	102	-0.124 (0.637)	-0.232 (0.031; 0.032)	-0.208 (0.025)
24553	102	-0.130 (0.625)	-0.091 (0.034; 0.035)	-0.125 (0.028)
51152	102	-0.134 (0.639)	-0.209 (0.030; 0.030)	-0.212 (0.025)

Health States	n	Observed Mean (SD)	Basic OLS Model Predicted Value (SE; $SE_{Cluster}$)	Basic RE Model Predicted Value (SE_{RE})
11122	99	0.802 (0.326)	0.684 (0.023; 0.020)	0.677 (0.021)
35311	99	0.263 (0.579)	0.222 (0.028; 0.030)	0.207 (0.024)
42321	99	0.219 (0.652)	0.246 (0.035; 0.038)	0.227 (0.028)
13224	99	0.207 (0.654)	0.293 (0.032; 0.030)	0.300 (0.026)
34232	99	0.152 (0.624)	0.115 (0.032; 0.035)	0.122 (0.026)
52335	99	-0.174 (0.631)	-0.191 (0.029; 0.032)	-0.180 (0.024)
24445	99	-0.345 (0.601)	-0.342 (0.030; 0.030)	-0.326 (0.025)
43555	99	-0.372 (0.583)	-0.404 (0.029; 0.031)	-0.416 (0.024)
21334	98	0.189 (0.638)	0.114 (0.028; 0.031)	0.116 (0.024)
32314	98	0.163 (0.671)	0.160 (0.030; 0.033)	0.154 (0.025)
12334	98	0.152 (0.646)	0.192 (0.028; 0.033)	0.151 (0.024)
23242	98	0.083 (0.675)	0.132 (0.030; 0.035)	0.115 (0.025)
53412	98	0.021 (0.679)	-0.014 (0.028; 0.035)	-0.024 (0.024)
33253	98	-0.017 (0.661)	0.072 (0.032; 0.029)	0.065 (0.026)
24342	98	-0.026 (0.654)	-0.027 (0.035; 0.040)	-0.052 (0.028)
55225	98	-0.227 (0.658)	-0.200 (0.026; 0.027)	-0.200 (0.023)
11221	96	0.731 (0.391)	0.687 (0.024; 0.021)	0.692 (0.022)
11421	96	0.496 (0.536)	0.529 (0.026; 0.022)	0.543 (0.023)
13313	96	0.423 (0.574)	0.363 (0.033; 0.037)	0.367 (0.027)
25122	96	0.248 (0.662)	0.290 (0.031; 0.032)	0.315 (0.026)
11235	96	0.159 (0.670)	0.343 (0.028; 0.030)	0.314 (0.024)
12514	96	0.088 (0.667)	0.068 (0.027; 0.030)	0.052 (0.024)
31525	96	0.069 (0.644)	0.227 (0.024; 0.022)	0.222 (0.022)
12244	96	-0.012 (0.680)	0.032 (0.031; 0.036)	0.070 (0.026)
54231	96	-0.086 (0.680)	0.134 (0.030; 0.026)	0.129 (0.025)
45233	96	-0.111 (0.666)	-0.025 (0.030; 0.026)	-0.013 (0.025)
51451	96	-0.195 (0.644)	-0.083 (0.030; 0.034)	-0.087 (0.025)
34515	96	-0.200 (0.658)	-0.082 (0.030; 0.033)	-0.055 (0.025)
55233	96	-0.215 (0.667)	-0.255 (0.033; 0.037)	-0.266 (0.027)
35245	96	-0.232 (0.636)	-0.123 (0.027; 0.033)	-0.117 (0.024)
45144	96	-0.303 (0.632)	-0.221 (0.029; 0.035)	-0.226 (0.025)
52455	96	-0.344 (0.630)	-0.467 (0.028; 0.033)	-0.466 (0.024)
11212	90	0.746 (0.428)	0.675 (0.027; 0.023)	0.673 (0.023)
12112	90	0.701 (0.466)	0.638 (0.024; 0.023)	0.632 (0.022)
23152	90	0.054 (0.698)	0.200 (0.030; 0.035)	0.168 (0.026)
21345	90	-0.019 (0.670)	0.034 (0.028; 0.026)	0.027 (0.024)
43514	90	-0.038 (0.680)	-0.228 (0.028; 0.036)	-0.227 (0.024)
34244	90	-0.092 (0.687)	-0.060 (0.028; 0.028)	-0.076 (0.024)
55424	90	-0.248 (0.681)	-0.346 (0.032; 0.035)	-0.351 (0.026)
44553	90	-0.270 (0.673)	-0.390 (0.031; 0.036)	-0.388 (0.026)

EQ-5D-5L: 5-Level EuroQol 5-Dimension; n: Number of participants elicited the given health state; SD: Standard deviation; OLS: Ordinary least square regression model; SE: Ordinary least square standard error; $SE_{Cluster}$: Cluster-robust standard error; RE: Random effect model; SE_{RE}: Random effect standard error.