MAPPING THE EQ-5D-3L UTILITY SCORE FROM PEDSQL™ 4.0 GENERIC CORE SCALE IN PEDIATRIC TRANSFUSION-DEPENDENT THALASSEMIA IN MALAYSIA

IRWINDER KAUR CHHABRA A/P NARINDAR

UNIVERSITI SAINS MALAYSIA

2020

MAPPING THE EQ-5D-3L UTILITY SCORE FROM PEDSQL™ 4.0 GENERIC CORE SCALE IN PEDIATRIC TRANSFUSION-DEPENDENT THALASSEMIA IN MALAYSIA

by

IRWINDER KAUR CHHABRA A/P NARINDAR

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

December 2020

ACKNOWLEDGEMENT

The completion of this thesis would not have been possible if it were not for the support that I have received from the wonderful people around me.

First and foremost, my deep gratitude goes to Associate Professor Dr Asrul Akmal Shafie for taking me under his wings. Thank you for the opportunity to learn under your guidance. Your passion, guidance and expertise in the subject matter never ceases to inspire me to push further for improvement.

My appreciation and love extend to my fellow colleagues in the Discipline of Social and Administrative Pharmacy, Annushiah Vasan Thakumar, Haarathi Chandriah, Jacqueline Wong Hui Yi, Jovelyn Tan Bee Ying, Khairu Hazwan, Ng Chin Hui, Noor Syahireen Mohammed, Siti Fauziah binti Abu and Sivaraj Raman. Your friendship, stimulating discussions on our research and the experiences that we shared throughout these two years is something I will treasure for life. My gratitude also goes out to the staff at Universiti Sains Malaysia's School of Pharmaceutical Sciences, particularly Mr Mohammed Firdauz bin Isahak, Mr Ahmad Faiz bin Badiozaman and Cik Nurul Jannah Kamaruz'zaman. Thank you for your efficiency and help in navigating the bureaucracy of the university system.

I would also like to acknowledge the Ministry of Health, the paediatricians, the interviewers, transfusion-dependent thalassemia patients, parents and caregivers who willingly participated in this research. We hope that our research findings would ultimately serve to improve how you experience healthcare in our country.

Last but not least, I would like to thank my husband, Harrison WHD, my family and friends for their patience, encouragement, and support throughout this journey. I would not have made it without you.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

AIC	Aikake Information Criteria
BIC	Bayesian Information Criteria
CEA	Cost effectiveness analysis
CHU9D	Child Health Utility Index – 9 dimensions
CLAD	Censored least absolute deviations
DFO	Desferrioxamine
DFP	Deferiprone
DFX	Deferasirox
DMARD	Disease modifying antirheumatic drugs
EFS	Emotional Functioning Score
GDP	Gross domestic product
GLM	Generalized linear model
HAQ	Health Assessment Questionnaire
HRQoL	Health-related quality of life
HSUV	Health state utility values
HUI 3	Health Utility Index 3
ICC	Intraclass correlation coefficients
ICER	Incremental cost effectiveness ratio
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
MAE	Mean absolute error
MLOGIT	Multinomial logistic regression
MREC	Medical Research and Ethics Committee
MSE	Mean squared error

NICE	National Institute for Health and Care Excellence
NMRR	National Medical Research Register
NPBM	Non preference-based measure
NTDT	Non-transfusion dependent thalassemia
OLOGIT	Ordinal logistic regression
OLS	Ordinary least squares
PBM	Preference based measure
PCHS	Psychosocial Health Summary Score
PedsQL 4.0 GCS	PedsQL 4.0 Generic Core Scales
PHSS	Physical Health Summary Score @ Physical Functioning Score
PO	Oral route
QALY	Quality adjusted life years
QoL	Quality of life
RM	Ringgit Malaysia
RMSE	Root mean squared error
SC	Subcutaneous route
ScFS	School Functioning Score
SD	Standard deviation
PedsQL 4.0 SF-15	Short Form 15 (PedsQL 4.0 Variant)
SF-12	12 Item Short Form Survey
SF-36	36 Item Short Form Survey
SF-6D	Short Form-6 Dimension
SFS	Social Functioning Score
TDT	Transfusion dependent thalassemia

TPM	Two-part model
TSS	Total Summary Score
UK	United Kingdom
US	United States
USD	United States Dollars
VAS	Visual Analogue Scale
WHOQOL-BREF	World Health Organization Quality of Life Instrument

LIST OF APPENDICES

PEMETAAN SKOR UTILITI EQ-5D-3L DARI PEDSQL 4.0 GENERIC CORE SCALE DALAM PESAKIT PEDIATRIK TALASSEMIA BERGANTUNG TRANSFUSI DI MALAYSIA

ABSTRAK

Sebaik-baiknya, ujian klinikal yang mengukur keberkesanan rawatan harus merangkumi skala ukuran berdasarkan pilihan (*preference-based measure*) yang dapat menghasilkan nilai utiliti kesihatan untuk digunakan dalam analisa keberkesanan kos. Walaubagaimanapun, kebanyakan ujian klinikal tidak menggunakan skala ukuran tersebut, menjadikannya sukar untuk melaksanakan penilaian ekonomi. Pemetaan merupakan satu kaedah untuk menganggarkan hubungan statistik antara skala bukan berdasarkan pilihan (*non-preference-based* measure) dan skala berdasarkan pilihan. Objektif kajian ini adalah untuk memetakan skala bukan berdasarkan pilihan *PedsQL Generic Core Scales(GCS)* ke skor utiliti EQ-5D-3L dalam populasi pesakit talasemia pediatrik yang bergantung kepada transfusi darah (*transfusion-dependent Thalassemia*) dan menentukan kesan penggunaan nilai utiliti tersebut dalam analisa keberkesanan kos terapi pengelatan besi. Kajian ini menggunakan tinjauan keratan rentas. Sebanyak 368 pesakit mengambil bahagian di dalam tinjauan ini dan maklumat dikumpulkan menggunakan borang pengumpulan data, skala *PedsQL GCS* dan EQ-5D-3L yang sesuai dengan usia. Menggunakan skala *PedsQL GCS*, nilai skor min Ringkasan Keseluruhan (*Total Summary Score*) (sisihan piawai) adalah 80.12 (13.87). Skor utiliti min (sisihan piawai) EQ-5D-3L adalah 0.909 (0.145) sementara skor min (sisihan piawai) EQ VAS adalah 81.22 (16.92). Skala EQ-5D-3L disahkan menggunakan skor skala *PedsQL* dan hipotesis apriori dari tinjauan kajian yang pernah dijalankan. Hasil daripada proses pengesahan ini, didapati bahawa skala EQ-5D-3L

dapat membezakan dan mengesan perubahan status kesihatan di kalangan populasi pesakit talasemia pediatrik. Enam kaedah regresi telah digunakan untuk menganggarkan algoritma pemetaan *PedsQL GCS* kepada EQ-5D-3L: *Ordinary Least Squares* (OLS), OLS dengan utiliti yang diubah log, model linear teritlak (*Generalized Linear Model* (GLM)), model dua bahagian (*Two-part Model* (TPM)), regresi *Tobit* dan regresi logistik multinomial (*Multinomial Logistic Regression* (MLOGIT)). Model terbaik dinyatakan dengan skor subskala *PedsQL GCS* dan dianggarkan menggunakan regresi GLM dengan pautan log dan pengedaran Gaussian. Ralat punca min (*Root Mean Square Error*) ialah 0.1301 dan ralat mutlak min (*Mean Absolute Error*) ialah 0.1037 menggunakan model tersebut. Walaubagaimanapun, proses pemetaan memperkenalkan ketidakpastian ke dalam nilai utiliti. Apabila utiliti yang dipetakan digunakan dalam analisa keberkesanan kos untuk rawatan pengelatan besi, nilai utiliti dan *Quality Adjusted Life Years(QALY)* yang lebih rendah dihasilkan, disamping nisbah keberkesanan kos tambahan (*Incremental Cost Effectiveness Ratio*) yang lebih tinggi. Kos per QALY menggunakan utiliti asal bernilai antara RM 263,294 hingga RM557,814 berbanding dengan kos per QALY RM2,365,583 apabila menggunakan utiliti dipetakan. Perbezaan besar antara hasil ini akan menghasilkan kesimpulan yang berbeza mengenai keberkesanan kos dan mempengaruhi akses pesakit kepada rawatan. Kesimpulannya, walaupun pemetaan dapat menghasilkan nilai utiliti, pengumpulan data utiliti secara langsung masih dianggap penyelesaian yang terbaik.

MAPPING THE EQ-5D-3L UTILITY SCORE FROM PEDSQL 4.0 GENERIC CORE SCALE IN PEDIATRIC TRANSFUSION-DEPENDENT THALASSEMIA IN MALAYSIA

ABSTRACT

Ideally, clinical trials measuring the effectiveness of an intervention should include a preference-based measure which can generate health state utility values. Utility values are key drivers of cost effectiveness analysis. However, most trials use non-preference-based instruments, making it difficult to perform economic evaluations. Mapping is a process of predicting a statistical relationship between a non-preference-based and a preference-based measure. The objective of this study was to map the non-preference based PedsQL Generic Core Scales instrument onto the EQ-5D-3L utility score in a population of paediatric transfusion-dependent thalassemia(TDT) patients and determine the impact of using these mapped utility values on the cost effectiveness outcome of iron chelating therapy. Using a cross-sectional survey and a multi-stage sampling, a total of 368 patients were sampled using data collection forms, age appropriate PedsQL GCS and EQ-5D-3L instrument. On the PedsQL GCS, the Total Summary Score mean (standard deviation) was 80.12 (13.87). The EQ-5D-3L utility value mean (standard deviation) was 0.909 (0.145) while the EQ VAS mean (standard deviation) score was 81.22 (16.92). The EQ-5D-3L instrument was validated against the PedsQL GCS and a priori hypothesis derived from literature. The instrument demonstrated convergent, discriminant and known-group validity, indicating its ability to differentiate and recognize health state changes. Six regression methods were used to predict the mapping algorithm: the ordinary least squares (OLS), OLS with log transformed utilities, generalized linear model (GLM),

two-part model (TPM), Tobit and multinomial logistic regression (MLOGIT). The model with the best predictive ability was specified by the PedsQL GCS subscale scores, estimated using a Gaussian distribution GLM and a log link. The root mean square error (RMSE) was 0.1301 and the mean absolute error (MAE) was 0.1037. However, the process of mapping introduced uncertainty into utility values. When mapped utilities were used in a cost effectiveness analysis for the treatment of iron overload complications in TDT, the mapped utilities generated lower utility values, underestimating the quality-adjusted life year (QALY) gains, resulting in higher incremental cost-effectiveness ratios (ICERS). The cost per QALY using the observed utility ranged from RM 263,294 to RM557,814 compared to the RM2,365,583 per QALY gain when using mapped utilities. This vast difference between outcomes could result in different conclusions about the cost effectiveness analysis and affect patients access to treatment. In conclusion, although mapping can facilitate the generation of utility values, directly collected utility data is still considered the best option.

CHAPTER 1 INTRODUCTION

The advancement of health technologies and interventions has increased the demand for health care services and available resources. The systematic identification of the cost and consequences of health intervention options aids in determining which option maximizes the value of health care spending. This process, called an economic evaluation, can be used to prioritize resource allocation(Drummond, 2017). Cost-effectiveness analysis (CEA) is one form of economic evaluation which involves the calculation of monetary costs and the Quality Adjusted Life Years (QALYs) as an outcome of intervention. Although health outcomes can be measured in other natural health units such as pain-free days and survival rates, using the QALY enables the comparison of outcome across different health care areas(Whitehead & Ali, 2010).

A QALY is able to reflect health-related quality of life and survival in a single index and is computed by multiplying a utility value of a defined health state with the duration of time spent in the health state. Health-state utility values reflect an individual's preference for different health outcomes and is measured on a scale of 0 to 1. A health state of 0 represents death and 1 represents a perfect health state(Rascati, 2009). States worse than death can also be accounted for when the states take a negative value(Whitehead & Ali, 2010). Examples of health states worse than dead has been associated with extremely low quality of life(Bernfort, Gerdle, Husberg, & Levin, 2018) and also in certain cancer stages(Chou, Chiang, & Ko, 2020).

Instruments that measure health state outcomes can be divided into preference-based or non-preference-based measures as illustrated in Figure 1.1. The main difference lies in the ability of the measure to produce health state utility values which has a weighted

score or ‘value’ attached to it. Value is measured in terms of ‘desirability’ or preference for the described health state.

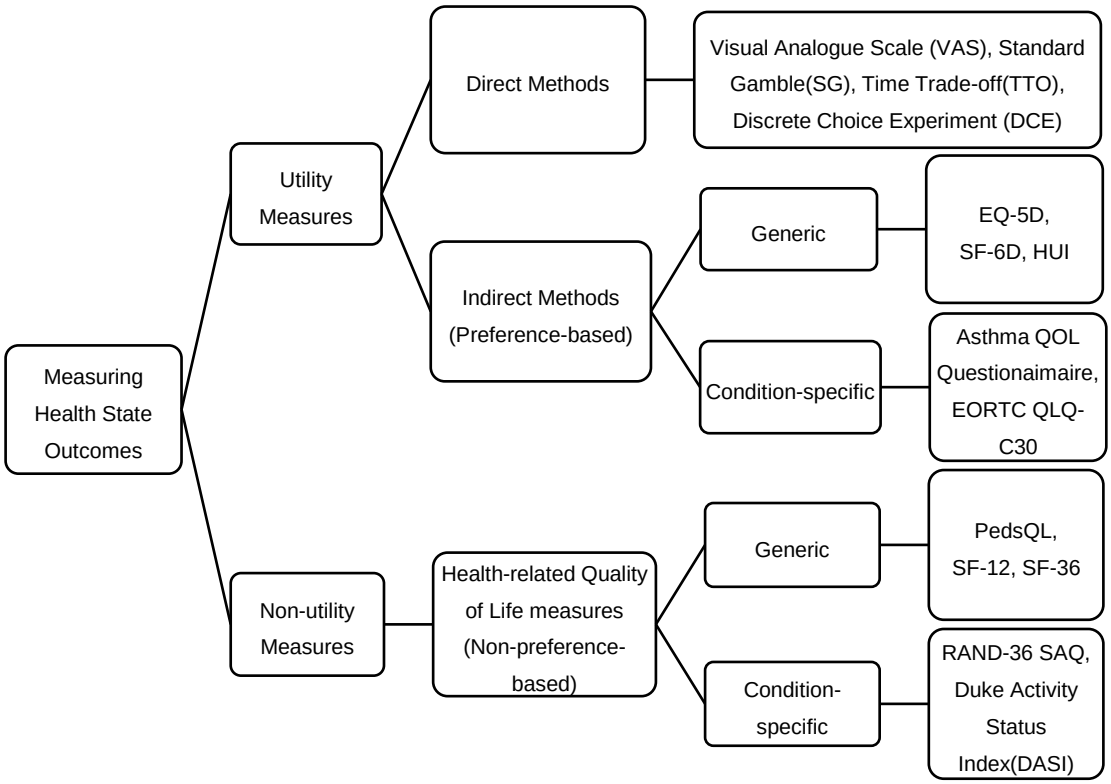

Figure 1. 1 Classification of instruments that measure health state outcomes

An example of a non-preference-based measure (NPBM) is the generic PedsQL Generic Core Scales (GCS) as illustrated in Figure 1.2 below. The measure has a descriptive component, covering both physical and psychosocial dimensions. On the PedsQL, the score ranges from 0 to 100, with higher scores indicating a better quality

of life. The measure describes the patient's perception of his or her condition without a valuation of preference.

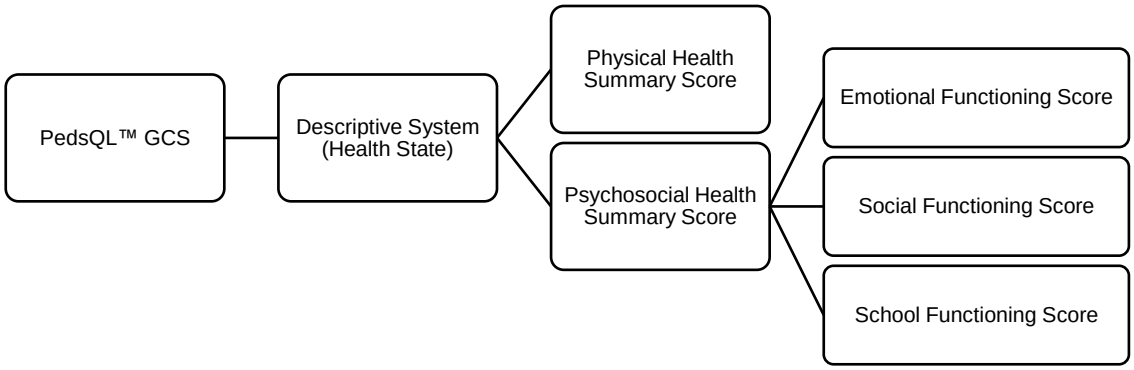

Figure 1. 2 Components of the NPBM PedsQL Generic Core Scale instrument

Similarly, a generic utility measure such as EuroQol's EQ-5D also has a descriptive component, covering domains such as the mobility, self-care, usual activity, pain/discomfort, and anxiety depression (Figure 1.3). However, in addition to that, the utility measure would have a value set which reflects a preference or weighted score obtained from the general population. This preference valuation for health states enables the generation of utility values.

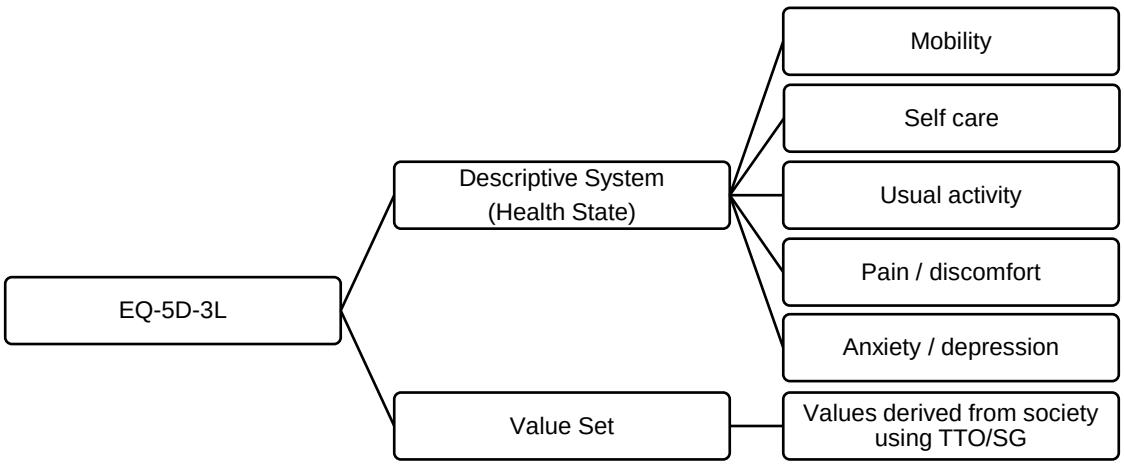

Figure 1. 3 Components of the PBM EuroQol's EQ-5D instrument

Utility values can be measured by first defining the health states of interest and then valuing those health states using methods such as the visual analogue scale (VAS), standard gamble (SG), time trade-off(TTO) or the discrete choice experiment(DCE)(Rascati, 2009). Alternatively, utility values can be measured indirectly using a pre-scored preference-based questionnaire. Generic questionnaires such as the EQ-5D, SF-6D or the Health Utility Index (HUI) covers the general aspects of health such as the physical, emotional and social dimensions, allowing it to be utilized and compared across various conditions. However, generic measures may be less sensitive to particular aspects of a condition, leading to the development of condition-specific measures such as the Asthma Quality of Life (QoL) Questionnaire and the European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30).

Thalassemia is an inherited blood disorder which is defined by chronic anaemia. It is caused by the absence or decreased production of one or more of the four globin chains in a haemoglobin molecule. Depending on the need for blood transfusion, it can be divided into transfusion-dependent thalassemia (TDT) or non-transfusion-dependent thalassemia (NTDT) (M.-D. Cappellini, Cohen, Porter, Taher, & Viprakasit, 2014b). According to the recent 2018 Malaysian Thalassemia Registry report(Ibrahim, 2019), there are approximately 7,984 thalassemia patients in the country.

Severe anaemia in TDT is an indication for regular blood transfusions, failing which patients may experience growth retardation and or decreased survival. However, regular blood transfusions in turn causes iron overload in the major body organs as the

body lacks a mechanism to excrete these excess irons. The toxic accumulation of iron in the major organs of the body often result in heart failure, cirrhosis, growth retardation, endocrine abnormalities and other complications(*Management of Transfusion Dependent Thalassaemia*, 2009). There are currently three iron chelating agents available for the treatment of iron overload: a subcutaneous injection Desferrioxamine (DFO) and two other oral agents, Deferiprone (DFP) and Deferasirox (DFX). Although desferrioxamine is the recommended first line drug, oral Deferasirox, which was introduced in 2010, has been recommended to help improve adherence to treatment. However, the cost of Deferasirox is significantly higher as compared to Desferrioxamine.

Tables 1.1 and 1.2 below summarize the clinical characteristics and the side effects of the various iron chelators(M.-D. Cappellini et al., 2014b):

Table 1.1 Clinical characteristics of iron chelating therapy

Characteristics	**Desferrioxamine (DFO)**	**Deferiprone (DFP)**	**Deferasirox (DFX)**
Dose range (mg/kg/day)	Pediatric: 20 - 40 Adult: Up to 50 – 60 (Infuse over 8 - 24 hours for 3-7 times per week)[3]	75 - 100 (in 3 divided doses)	20 - 40 (once daily dosing)
Suitable Age Group i) Age 2-6	First line treatment (Caution in children < 3 years old in view potential toxicity)	Not recommended (Insufficient information for licensing)	First line treatment
ii) Age > 6 and adults	First line treatment	If other chelation or DFO not tolerated or ineffective	First line treatment
Dosage Form	500mg vial	500mg tablet	125mg & 500mg dispersible tablet
Brand Name	Desferal	GPO-L1 Ferriprox Kelfer	Exjade
Route of Administration	Parenteral (Slow SC or IV infusion)	Oral	Oral
Half-life	20-30 minutes	3-4 hours	12-16 hours
Iron excretion	Both urine & faeces	Urine	Faeces
Chelation efficiency (% of drug excreted iron)	13	7	27

Table 1.2 A summary of side effects related to iron chelating therapy

Side Effects	Desferrioxamine (DFO)	Deferiprone (DFP)	Deferasirox (DFX)
Skin - local pain, infection at injection sites	6 - 85%	Not Applicable	Not Applicable
Skin rashes	As Above	Dry skin if associated with zinc deficiency	7%
Hematological	Not Available	Agranulocytosis (ANC < 500 x 103/L) & Neutropenia (ANC < 1500 x 103/L) Thrombocytopenia in 45% of children < 7 years old.	Not Available
Gastro-intestinal symptoms	24%	33%	15.2% raised liver enzymes
Renal (↑ SCr)	14%	Not Available	38% .
Joint pain & stiffness	18 - 19%	28 - 37.5%	Not Available
Reduced visual acuity & impaired visual field	Present if dose is high.	Not Available	< 1% lens opacities.
Sensorineural deafness	Present if dose is high	Not Available	< 1% high frequency hearing losses.

1.1 Problem statement

The number of thalassemia patients in the country are hypothesized to increase over the years as approximately 4.5% of Malaysians are heterozygous carriers of beta thalassemia and are at risk of having a child affected by the condition(George, 2001). This is expected to put a substantial burden on the healthcare system as most of the treatments are subsidized by the Ministry of Health. Conducting economic evaluations to determine whether a drug is cost-effective in the long run would ensure that resources are spent efficiently.

Ideally, clinical trials that measure the effectiveness of an intervention should include a preference-based measure which can generate health state utility values for use in the generation of QALYs and the CEA. A literature search revealed that most of the effectiveness data pertaining to iron chelating agents were collected using non-preference-based instruments, making it difficult to perform economic evaluations. This is because non-preference-based instruments does not incorporate preferences onto its health states, making it impossible to generate utility values which are required in economic evaluations.

For example, although the Short Form 36 (SF-36) has the ability to detect health changes in a population following an intervention, the scoring algorithm assumes equal interval between the response choices, without a desirability value(preference) attached to the responses. A preference-based measure like the Short Form 6D (SF-6D) on the other hand is a questionnaire with a scoring function to weight the responses according to preferences for certain health condition over others. To bridge this gap between NPBM and PBM, a method called mapping has been introduced.

Mapping is a method of predicting a statistical relationship between a non-preference-based and a preference-based measure. The resulting algorithm or model is then used to predict the utility values of the preference-based outcomes from the available HRQoL or patient-reported outcomes. Subsequently, the predicted utility is used for the calculation of QALY's(L. Longworth & Rowen, 2013) in economic evaluations.

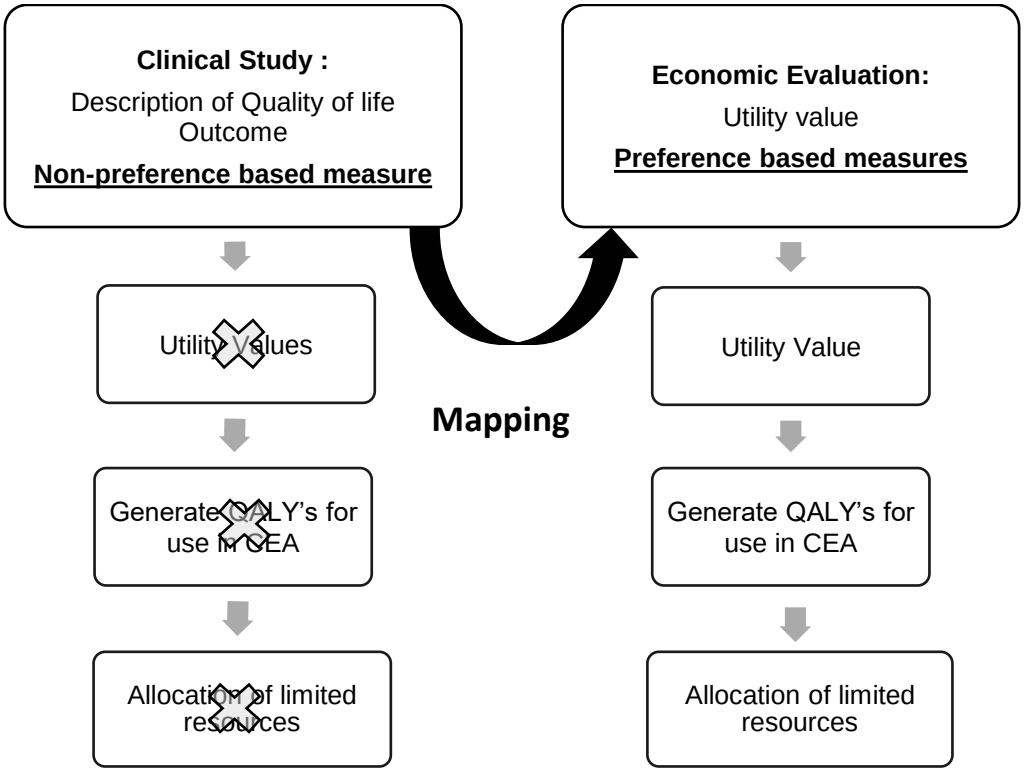

Figure 1. 4 An illustration of mapping

There were four main gaps in knowledge that was identified, and this research aims to fill those gaps.

1.1.1 Health-related quality of life of paediatric transfusion-dependent thalassemia patients in Malaysia

The last HRQoL survey of TDT patients in Malaysia was conducted in 2010. Most surveys were performed in urban areas with a limited sample size. The availability of iron chelation therapy was limited to desferrioxamine, hence a comprehensive

understanding regarding the effect iron chelation therapy such we deferiprone and deferasirox on HRQoL dimensions were lacking. The surveys were performed using non-preference-based instruments such as the PedsQL GCS and the SF-36. As of date, there is no study that describes the health state utility values for the transfusion-dependent thalassemia population in the country.

This research aims to update the HRQoL status of Malaysian children affected by TDT, with specific emphasis on the iron chelating therapy, using the PedsQL™ Generic Core Scales and EuroQol's EQ-5D-3L. In addition to describing the health profiles, this study aims to derive health state utility values which can be used in cost effectiveness analyses.

1.1.2 The validity of the EQ-5D-3L in the transfusion-dependent thalassemia population

Although the EQ-5D-3L has been validated for use in the Malaysian population(Faridah, Jamaiyah, Goh, & Soraya, 2010; Shafie, Hassali, & Liau, 2011), there is a gap in literature regarding the validity of its use in the thalassemia population, especially in children(Noyes & Edwards, 2011). It is pivotal that the instrument can measure, detect changes and differentiate between the severity of health states in this population before it is used as a basis for clinical or economic decision making(Fitzpatrick et al., 1992).

This research aims to examine the cross-sectional convergent and known-group validity of the EQ-5D-3L instrument in children with transfusion-dependent thalassemia.

1.1.3 Mapping the PedsQL™ Responses to Estimate EQ-5D-3L Values

A literature search revealed that most of the HRQoL data pertaining to TDT patients in the country was measured using the PedsQL GCS instrument. The data available could potentially be used for economic evaluations and the decision-making process. Mapping allows older studies which may not have included a preference-based measure to remain as part of the evidence base as comparators for economic evaluations of new interventions. The mapping approach has been endorsed by organizations such as The National Institute for Health and Care Excellence (NICE) for use when directly obtained utility data is not available(Louise Longworth & Rowen, 2011).

This study attempts to establish a mapping algorithm with the best predictive accuracy for converting PedsQL™ GCS onto the EQ-5D-3L using HRQoL data collected from transfusion-dependent thalassemia patients. This is the first study attempting to predict an algorithm between the PedsQL GCS and the EQ-5D-3L in this disease-specific population.

1.1.4 The impact of using mapped utilities on cost-effectiveness analysis

Although the same data is used in a mapping process, the use of different mapping algorithms is expected to produce different utilities. This may result in different cost – effectiveness results within the same economic evaluation. Hence, there is a need to understand the impact of using mapped utilities in a CEA compared to directly observed values.

The incremental cost-effectiveness ratio (ICER) is a statistic used in CEA to summarize the cost-effectiveness of an intervention. It represents the average

incremental cost associated with 1 additional unit measure of effect. The effect can be measured in units of QALY, making the ICER synonymous with the cost per QALY gained. The ICER can be used as a decision rule in resource allocation. If a decision maker is able to establish a willingness-to-pay value for the outcome of interest, this value can be adopted as a threshold value. If an intervention's ICER is above the threshold, it will be considered too expensive and would not be funded. Conversely, if the ICER's value is less than the threshold, it is considered cost-effective.

This study would examine the impact of using mapped utilities on a CEA of desferrioxamine and deferasirox in the treatment of iron overload complications in transfusion-dependent thalassemia, using QALY's as a measure of effect.

1.2Objective & scope of study

The main objective of this research was to map the PedsQL GCS values unto the EQ-5D-3L utility score and determine its impact on the cost effectiveness outcome of iron chelating therapy used in transfusion-dependent thalassemia.

The specific objectives are listed below:

1. To describe the PedsQL Generic Core Scales (GCS) and EQ-5D-3L profiles of thalassemia patients in Malaysia and its predictors.
2. To validate the EQ-5D-3L instrument among paediatric transfusion-dependent thalassemia patients in Malaysia.
3. To determine the best fitting statistical model to map the value of PedsQL GCS onto the EQ-5D-3L index
4. To determine the impact of using mapped utilities in the cost effectiveness analysis of iron overload treatment in transfusion-dependent thalassemia

CHAPTER 2 LITERATURE REVIEW

2.1 Health-related quality of life (HRQoL) of pediatric transfusion-dependent thalassemia patients

Studies have shown that a diagnosis of thalassemia affects a patient's quality of life(Telfer et al., 2005). Thalassemia may result in physical changes of the body. Inadequate blood transfusions are expected to result in thalassemia facies and stunted growth. This will affect a child's confidence and perception of themselves, causing them to have low self-esteem or feel stigmatized. Complications such as heart failure may limit the child's ability to function physically as a normal child would have. Frequent hospital visits for doctor appointments, blood transfusions and long hours of iron chelating therapy infusion is expected to disrupt a child's schooling attendance and social life. Patients have also reported that fatigue brought about by the anemia often results in a disrupted daily routine (Wahab et al., 2011).

A Health-Related Quality of Life (HRQoL) instrument measures the impact of a disease and its treatment on a patients physical and psychological wellbeing. HRQoL information can aid clinicians to comprehend a patients' needs beyond the clinical markers of the disease. Policy makers and program planners may also utilize the findings to guide and justify changes or improve the delivery of health care services to patients with the goal of improving their overall outcomes(Pennacchini, Bertolaso, Elvira, & De Marinis, 2011).

Health-related quality of life surveys in children with transfusion-dependent thalassemia have been conducted in Malaysia(A. Ismail, Campbell, Ibrahim, & Jones, 2006; M. Ismail et al., 2013; Sazlina, Asauji, & Juni, 2015), Thailand(Surapolchai P, 2010; Thavorncharoensap et al., 2010; Torcharus & Pankaew, 2011), India(Chordiya,

Katewa, Sharma, Deopa, & Katewa, 2018; Saha, Misra, & Saha, 2015; Shaligram, Girimaji, & Chaturvedi, 2007) and the Middle East(Abdul-Zahra, Hassan, & Ahmed, 2016; Ansari, Baghersalimi, Azarkeivan, Nojomi, & Rad, 2014; Arian, Mirmohammadkhani, Ghorbani, & Soleimani, 2019; Caocci et al., 2012; Meysam Seyedifar et al., 2016). The principles of treating TDT are similar across the regions, whereby most patients are treated with a combination of regular blood transfusions followed by the administration of iron chelating therapy to delay the development of iron overload complications. However, when the HRQoL outcomes in Malaysia are compared to the outcomes in neighbouring Thailand during the same period, the TSS score in Malaysia ranged between 65 to 70(A. Ismail et al., 2006; M. Ismail et al., 2013; Sazlina et al., 2015) while the TSS score in Thailand was higher with a range of 75 to 80(Surapolchai P, 2010; Thavorncharoensap et al., 2010; Torcharus & Pankaew, 2011). HRQoL outcomes are determined by individual experiences, beliefs, expectations, cultural differences and access to health care, explaining the differences in the HRQoL outcomes(Louis S Matza, Swensen, Flood, Secnik, & Leidy, 2004). This rationalizes the need for local studies to better understand the unique challenges faced by the patients in Malaysia.

A local survey comparing the HRQoL of thalassemia children with matched healthy controls using the PedsQL GCS instrument in 2005 found that thalassemia patients had a lower Total Summary Score(TSS) rating of 67.7 compared to the healthy controls who had a TSS of 79.51. (A. Ismail et al., 2006). This was followed by another study in 2009, comparing the scores of patients who self-reported against the report of parents who proxy-reported. It was found that patients self-report scores were lower (TSS = 65.35) compared to proxy report scores (TSS=67.20). This study highlights that self-report scores may not be directly comparable to proxy-reports.

Studies have shown that proxy-reported responses are not directly interchangeable with self-reports(Rand & Caiels, 2015) as proxy-reported surveys may either underestimate or overestimate HRQoL(Caocci et al., 2012; Chordiya et al., 2018; M. Ismail et al., 2013). However, there have been mixed findings regarding this issue. A study conducted on children with attention-deficit hyperactive disorder in the United States and the United Kingdom, showed that the proxy version of the EQ-5D-3L completed by parents were able to detect impairment in the children(Louis S. Matza, Secnik, Mannix, & Sallee, 2005).

There were also studies that examined the predictors of TDT's HRQoL. Significant predictors of a TDT patients HRQoL were pre-transfusion haemoglobin levels (M. Ismail et al., 2013), the presence of active treatment, side effects of iron chelating therapy and the duration of time since diagnosis(Sazlina et al., 2015). On the other hand, a caregiver's HRQoL were significantly associated with the number of thalassaemic children and the caregivers level of education (M. Ismail et al., 2013).

The availability of additional resources such as funds, clinical practice guidelines, new iron chelating therapy options, prevention and screening programs since 2009 has paved the path for a better quality of care in the treatment of TDT in Malaysia(Azman et al., 2016; *Management of Transfusion Dependent Thalassaemia*, 2009). The free supply of the oral iron chelators deferiprone and deferasirox to patients through public hospitals has increased the accessibility of treatment(Azman et al., 2016). These new initiatives from the Ministry of Health are predicted to improve the overall HRQoL of Thalassemia patients in the country.

2.2 The validity of the EQ-5D-3L instrument in measuring HRQoL

Validation of an instrument ensures that the scores obtained from an instrument is measuring the intended construct of HRQoL. A valid instrument would be able to recognize and differentiate between different health states, detect changes in health states and distinguish between different group of patients. Validity is not referring to the instrument itself, but rather to the interpretation of the instrument with a particular population group(Fayers & Machin, 2013). The construct validity EQ-5D-3L instrument has been validated among the Malaysian population in a cross-sectional study involving 596 Malaysian adults. The EQ-5D instrument was assessed against the SF-12 instrument along with several known demographic and illness characteristics of the participants(Shafie et al., 2011). However, there exists a knowledge gap on whether the EQ-5D instrument is valid within a population of thalassaemic children in Malaysia.

In a systematic review that describes the use of EQ-5D in children in 29 studies, it was noted that different EQ-5D versions have been used in children but information regarding its psychometric properties were limited. Four versions of the EQ-5D have been used, the EQ-5D-3L, EQ-5D-Y, Dutch EQ-5D child version and the EQ-5D+C, an extended questionnaire with a cognitive dimension. The EQ-5D-3L has been used in children aged 10 years and older, while the proxy version was completed by their caregivers in children who were younger. Both the EQ-5D-3L and EQ-5D-Y versions demonstrated ceiling effects, causing problems with sensitivity(Noyes & Edwards, 2011). Ceiling effects occur when a high proportion of respondents report the maximum score of the instrument, making the score distributions skewed and violating

statistical requirements of statistical analysis related to the prediction of preference-based scores or identification of predictor variables(Huang et al., 2008).

The EQ-5D-3L version was investigated for feasibility, construct, discriminant and criterion validity in children with orthopaedic problems. In orthopaedic children aged between 10 to 18 years old, feasibility was found to be greater than 98% in all domains. Construct validity of the EQ-5D-3L in relation to the SF-36 domains varied from 0.29 to 0.71, with the pain domain having the highest coefficient. The criterion validity which measured the correlation between the EQ-5D-3L score and the clinical severity score was 0.33. However, the EQ-5D-3L could not differentiate between groups of children with cerebral palsy and idiopathic scoliosis(Vitale et al., 2001). With regards to Thalassemia, only concordance between child and proxy-reported EQ-5D scores were available, whereby correlations were found to be higher for observable domains such as mobility, self-care and usual activities, and lower in unobservable domains such as pain/discomfort and anxiety depression(Shaligram et al., 2007).

There are challenges related to the assessment of the HRQoL in children. Children may be too young to have insight into the causes and treatment of a disease. They may interpret questions differently compared to an adult, especially if their understanding and depth in a language limits them from completing the survey as it was intended. A systematic review that aimed to identify available HRQoL instruments used in children identified problems that were associated with the measurement of QOL in children. These include the definition and measurement of QOL, limited disease specific measures, inconsistency between child and parent reports, limited measures of self-report for children, lack of precision regarding the domains of QOL and cultural appropriateness of instruments(Eiser & Morse, 2001). With these challenges, it is

important that an instrument be validated in children prior to its use to measure the QOL.

2.3 Mapping a non-preference-based measure onto the EQ-5D-3L

The use of cost-effectiveness analysis (CEA) has been increasing amongst many health technology assessment agencies to establish whether the cost of a new intervention justifies the health benefits it can offer. Potential health benefits can be measured in terms of Quality Adjusted Life Years (QALYs), calculated by multiplying the number of years spent in a health state with health state utility values (HSUVs). Preference-based measures (PBM) are utilized to measure these utility values during clinical trials. In cases where a trial has not included a PBM, mapping estimates a statistical relationship between a source non-preference-based measure (NPBM) and a target PBM. This statistical relationship is then used to estimate HSUVs for the patients in a trial as it would have been measured using a PBM(L. Longworth & Rowen, 2013).

The ISPOR Task Force on Good Practices for Outcomes Research made recommendations for statistical methods of mapping to estimate health utility from non-preference-based measures(A. J. Wailoo et al., 2017). The process of mapping requires that patient responses (initial estimation data set) be available on both the PBM and the NPBM. The variables such as the demographic and clinical characteristics of the estimation set should be similar to the 'target' sample where the mapping algorithm is intended to be used. For mapping to be successful, the task force also recommends that the concepts measured in the NPBM source measure should match that of the target PBM.

Next, it is important that the statistical distribution of the data be examined prior to choosing a statistical model for the mapping. The task force does not advocate a for a specific model type as the performance of models may vary according to the characteristics of the PBM, disease, patient population and the nature of the explanatory variables used. However, it is important that the selection of the model be made based on existing empirical evidence of good performance which accommodates the key features of the PBM. The specification of the model can take many forms. The most common model specification involves the preference-based index as the dependent variable and the item or domain scores of the non-preference-based measure as the independent variable. Sociodemographic factors may be added in as independent variables and has shown to improve the accuracy of the mapping function. However, inclusion of these variables should take into consideration if they are routinely collected during trials, else the algorithm generated would be impractical.

Upon estimation, the model performance in terms of goodness of fit and predictive ability should be reported. One aspect of model fit that is important is the extent to which modelled values coincide with the observed values. Validation of the mapping model is relevant although it is not routinely required. Guidelines recommend empirical validation, where the model is estimated in two data sets either from two separate studies (external validation) or from splitting a single data set (internal validation). Other internal validation methods may be used should there be insufficient sample size to perform data splitting.

Figure 2.1 summarizes the stages of mapping as outlined by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR):

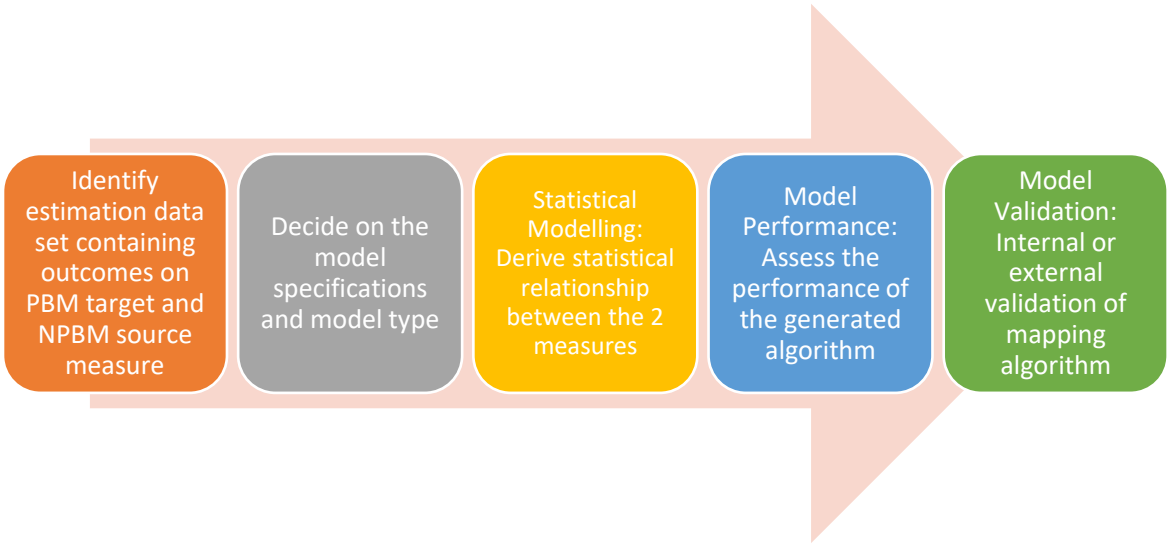

Figure 2. 1 Stages of mapping

Mapping methods can be divided into direct mapping or response mapping(Chuang & Whitehead, 2012). The direct mapping approach predicts the PBM utility scores directly from the NPBM using either the total or subdomain scores. The response mapping technique estimates the probability of obtaining a response level on the PBM based on the NPBM. Only then the predicted response levels are converted into a utility score(Hernández Alava, Wailoo, Wolfe, & Michaud, 2014). A systematic review conducted to identify mapping studies from January 2007 to January 2018 reported that out of the 233 mapping functions identified, a majority of the mapping functions (n=124) were mainly for the EQ-5D-3L(Mukuria et al., 2019).

The most common estimation method for mapping was the Ordinary Least Square (OLS) based on previous systematic reviews(Brazier, Yang, Tsuchiya, & Rowen, 2010; H Dakin, L Abel, R Burns, & Y Yang, 2018). This, however, does not necessarily mean that the OLS is the most appropriate method. One of the challenges

of modelling utility data lies in the distributional assumptions of the analysis methods(Pullenayegum et al., 2010). A common characteristic of health utilities is the non-normal distribution and the upper bound value of 1. In order to address this, researchers have shown a potential benefit in transforming health utility responses prior to fitting it with a linear regression model(Mitsakakis, Bremner, Tomlinson, & Krahn, 2020). Alternatively, other models such as the generalised linear model, Tobit and Censored Least Absolute Deviation (CLAD) have been explored to accommodate the non-normal distribution (Brazier et al., 2010; Mukuria et al., 2019).

A literature search on health-related quality of life of transfusion-dependent thalassemia patients revealed that most of the HRQoL instruments used were non-preference-based measures. Out of the 69 articles that were identified, 27 utilized the PedsQL GCS. Only three articles utilized the EQ-5D preference-based measure(Pratoombarn et al., 2015; M. Seyedifar et al., 2016; Shaligram et al., 2007). Most HRQoL surveys would utilize non-preference-based measures as it has the advantage of describing the quality of life and also the dimensions of interest in detail as compared to preference-based measures. Preference-based measures like the EQ-5D are brief (5 questions) and may not be able to capture all the dimensions of a disease as accurately. PBM are usually brief as valuation of health states are usually tedious and may cause fatigue to the respondents. In addition to that, the objective of PBMs are to generate health state utility values as compared to the objective of NPBM, which is to describe the different dimensions of a patient's health(Drummond, 2017).

To bridge this gap between the PedsQL and EQ-5D-3L, a mapping algorithm to convert between the two types of measures can be developed. Several studies mapping the PedsQL™ GCS were identified. Based on a general child population aged 11 to 15

years, Khan et. al. attempted to map PedsQL domain scores to the EQ-5D-Y utilities using the UK Adult tariffs. It was reported that the Ordinary Least Squares (OLS) method using PedsQL domain scores with their squared terms predicted the health utilities the best model(K. A. Khan, S. Petrou, O. Rivero-Arias, S. J. Walters, & S. E. Boyle, 2014). Other studies include the prediction of HUI3 utilities using the OLS in children with autism spectrum disorders(Payakachat et al., 2014) and CHU9D (Lambe et al., 2018; Mpundu-Kaambwa et al., 2017a; Sweeney, Chen, Gold, Mensah, & Wake, 2019), with mixed findings on the method with the best predictive accuracy. The available evidence regarding the mapping of PedsQL GCS and PedsQL Short Form-15 (SF-15) is summarized in Table 2.1.

Table 2. 1 Summary of evidence for PedsQL GCS Mapping

Article Title	Target Measure	Patient group and sample size (n)	Mapping models investigated	Summary Outcomes
Mapping EQ-5D Utility Scores from the PedsQL Generic Core Scales (K. A. Khan et al., 2014)	EQ-5D-Y	General population of children aged 11-15 years old (n=559)	OLS; GLM; 2-part; CLAD; Tobit; response mapping	The ordinary least squares (OLS) models with PedsQL™ GCS subscale scores, squared terms and interactions had the best prediction accuracy. Due to incomplete data on some of the response levels, the response mapping models could not be estimated.
Predicting health utilities for children with autism spectrum disorders (Payakachat et al., 2014)	Health Utilities Index 3(HUI-3)	Children aged 4-17 years old with autism spectrum disorders (n=224)	OLS only. The distribution of HUI3 scores were negatively skewed but did not have a ceiling effect and would not benefit from other regression methods.	Three source measures were used – the Child Behavior Checklist, Vineland-II and the PedsQL). Nine mapping algorithms were developed using the OLS with the source measures, the Autism Diagnostic Observation Schedule, child age and cognitive ability as independent predictors.

Mapping the Pediatric Quality of Life Inventory (PedsQL™) Generic Core Scales onto the Child Health Utility Index–9 Dimension (CHU-9D) Score for Economic Evaluation in Children(Lambe et al., 2018).	Child Health Utility Index–9 Dimension (CHU-9D)	Children with corticosteroid-sensitive nephrotic syndrome aged 5 years and above (n=279)	OLS, Tobit, GLM with 6 different sets of covariates	The generalized linear model (mean absolute error = 0.0408; mean square error = 0.0035) was the best performing model. The proportion of index scores deviating from the observed scores by < 0.03 was 53%.
Mapping CHU9D Utility Scores from the PedsQL™ 4.0 SF-15 (Mpundu-Kaambwa et al., 2017b).	Child Health Utility Index–9 Dimension (CHU-9D)	General population of adolescents aged 15-17 years old (n=755)	OLS, GLM, robust mixed model, multivariate factorial polynomial, beta-binomial, finite mixture model and multinomial logistic model.	The mixed models' estimator with selected PedsQL dimension scores had the best predictive accuracy using the mean absolute error and the equivalent beta-binomial model had the best predictive accuracy using mean squared error.