The effect of comorbidity on Health-Related Quality of Life utility values in a US representative survey: comparison of five models

By Hiba Saleh

The faculty of Health, Policy & Law
Erasmus University, Rotterdam
In partial Fulfilment of the Requirements of the Degree of
Master of Science

THESIS

Student number:	464464
Name supervisor:	Erwin Birnie, PhD
Course:	Master Thesis Health Economics, Policy and Law
Course code:	GW4555M
Word count:	10,286

5th of August, 2021

Table of Contents

Abstract

INTRODUCTION

Health utility scores quantify health-related quality of life (HR-QOL) along a continuum that typically ranges from 0.0 (dead) to 1.0 (full health), and are essential in developing summary measures of population health, as well as performing cost-effectiveness analysis (CEA) or cost-utility analysis (CUA) of different treatment and intervention strategies. A key methodological issue is that traditionally, utility scores have been developed primarily for single health conditions, even though comorbidities are common in both general and patient populations. Inaccuracies in health measurement are likely to occur when comorbidity is ignored in the estimation of utility scores. The multiplicative, additive, minimum, maximum and average models are mathematical models for combining single-condition utility scores. The empirical evidence for their performance is mixed, as well as difficult to assess due to a lack of standardization in utility instrumentation and analytical procedures used. The need for this specific research is evident in the current limitation of quantitative data regarding the estimation of comorbidity on health state utilities and accurately predicting the health state utilities. The present thesis has two main objectives. Firstly, to assesses the impact of comorbidity on EQ-5D health state utilities and secondly to choose the best performing model in predicting the actual EQ-5D health state utility values using goodness of fit parameters

METHODS

Data in this study was gathered using the Medical Expenditure Panel Survey (MEPS) 2002 of (n=39,165) individuals from the general population, of which (n=27,283) were 18 years or older with at least one comorbidity (n=7,506). Unadjusted and adjusted OLS linear regression models were used to compare EQ-5D health state utilities generated using the 5 models. The dependent variable was EQ5D health utility score computed according to the D1 model and the independent variables were age classes, socioeconomic status (SES) classes and chronic health conditions (diabetes, asthma, arthritis, stroke, joint pain, emphysema and high blood pressure). The goodness of fit was assessed in terms of the Adjusted R-Squared (R2), Mean Square Error (MSE) and 95% Confidence Interval (CI).

RESULTS

The multiplicative produced the highest adjusted R2(0.131) and lowest MSE (0.035) in the unadjusted analysis. Similarly, the multiplicative model produced the highest adjusted R2 (0.198) and lowest MSE (0.033) in the adjusted analysis. The maximum model performed very poorly, with an adjusted R2 (0.00) and MSE (0.041) in the unadjusted analysis and R2 (0.162) and MSE (0.041) in the adjusted analysis. All models showed a higher adjusted R2 in the adjusted analysis in comparison with the unadjusted analysis.

CONCLUSION

The multiplicative model had more accurate results than the other models. However, all five models showed a bad performance in predicting the actual EQ-5D health state utilities. Including more independent variables, using other parameters and other models (the hybrid mathematical model) might give better results. Further research is recommended to verify and specify the findings of this study.

1. Introduction

Guidelines on cost-effectiveness analysis (CEA) in the United States (US) and Europe suggest that the QALY method is the most appropriate measure of health effectiveness because of its ability to incorporate both survival (length of life) and the impact on health-related quality of life (HRQOL) associated with different health states into a single index. Utility scores (e.g., EQ-5D index) are used to calculate the number of quality-adjusted-life-years (QALYs) as a measure of health effectiveness in clinical and population studies and in CEA and cost-utility analyses (CUA) (W. Sullivan & V.H. Ghushchyan, 2012). Many studies designed to measure the impact of a particular disease or intervention on utility scores overlook the influence of comorbidity on utility scores. It is common to find analyses that adjust for demographic characteristics (e.g., age and gender) and even socioeconomic attributes (e.g., income and education) (W. Sullivan et al., 2012), but often the impact of comorbidity is omitted. The importance of comorbidity burden on utility scores is not well understood [ref], while at the same time, there is an increasing prevalence of comorbidities due to aging populations [ref].

1.1. Relevance

Comorbidity is a significant and growing problem that increases with age. It is common in both clinical and general populations (A. M. Broemeling et al., 2005; A.Z. Gijsen et al., 2001). Particularly in individuals with chronic conditions, comorbidity appears to be the norm rather than the exception. For example, a recent study of the British Columbia Linked Health Database (A. M. Broemeling et al., 2005) showed that among individuals (aged≥18 years) diagnosed with diabetes, many also had one or more chronic conditions. In 1999, Wu and Green estimated that 65% of individuals older than 65 years and 85% of individuals older than 85 had multiple chronic medical conditions. Individuals aged between 60 and 79 years have been estimated to have an average of 2.6 coexisting medical conditions, while those who are 80 years or older have an average of 3.6. A study from the RAND Health in the US showed that the number of Americans with comorbid chronic conditions would increase from an estimated 57 million in 2000 to 81 million in 2020 (F. Wu et al., 2000). In 2020, the number of individuals in the US with two or more chronic conditions was estimated to reach 81 million. This expected increase corresponds to the continued growth of elderly populations, in which comorbidities are particularly common (Boyd et al., 2005). With the increasing prevalence of comorbidity and multimorbidity in elderly, the rising cost of providing care for this population and the necessity to demonstrate the value of medical interventions, it is essential to understand the impact of chronic comorbidity on health utilities. This substantial social and economic toll of comorbidities makes it a high priority for clinicians, researchers, and health policy makers (A.Z. Gijsen et al., 2001).

Standard catalogs of utility scores have been developed to facilitate clinical decision analyses or treatment comparisons (Z. F. Alex, 2008). Although comorbidity is common in both general and patient populations (J. Hanmer, 2010), these utility scores have been developed primarily for single health conditions. Ignoring comorbidity in the estimation of utility scores can lead to inaccurate and biased utility scores that could impact the outcomes of CEAs, CUAs, decision analyses or burden of disease studies. For instance, in many CUAs it has been assumed that the baseline HR-QOL score for people with a certain health condition is exactly the reduction from perfect health (1.0 in the utility scale) resulting from that condition only, and that alleviation of that condition will restore affected persons to perfect health. However, due to possible comorbidities and other factors such as aging, many people with a given condition would not be in perfect health without that condition. Ignoring comorbidities in CUA can therefore exaggerate treatment benefits, threatening the generalizability of the results to real-life clinical settings where patients often have multiple conditions. A more realistic assumption is that a given treatment can, at best, raise the average HR-QOL for persons with the targeted condition to the same level as that of persons without that particular condition. Thus, when executing CUAs, it is important to accurately account for comorbidity since the results are intended to inform resource allocation in the health care sector. Moreover, this can enhance the evidence base necessary for health care decision-making (C.N. McIntosh, 2010).

Figure 1: *Idealized and actual HR-QOL benefits of a hypothetical new treatment for congestive heart failure (CHF), in persons 45 years and older (Feeny et al, 2002).*

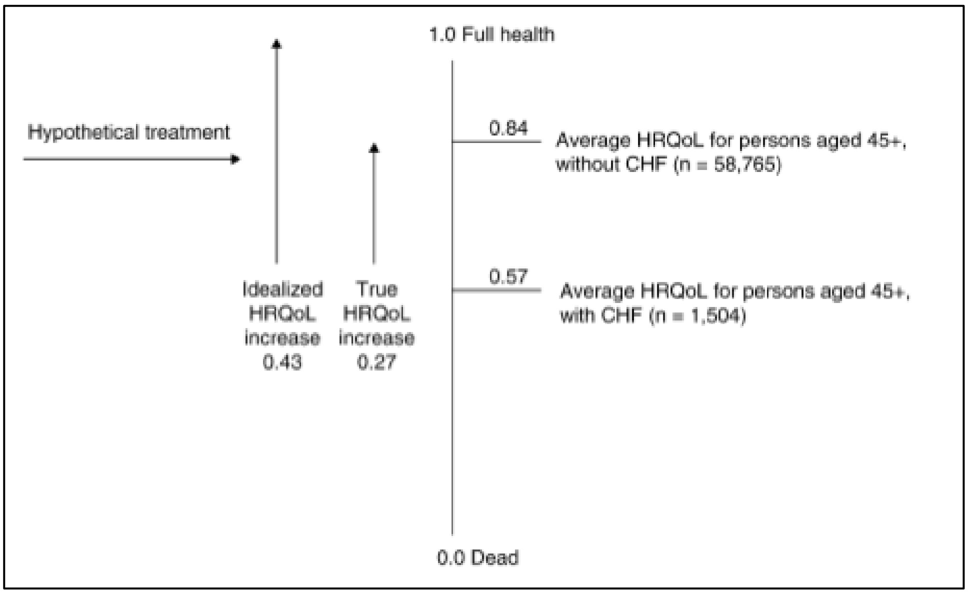


Figure 1 illustrates how ignoring the impact of comorbidity can artificially enhance the estimated benefits of a treatment. It is shown that persons aged 45 years and older without CHF are not in full health. But if one assumes that the cure of CHF will restore full health, the treatment specific HR-QOL gain could be overestimated by 0.16 points (0.43 – 0.27) on the utility scale.

At present, three approaches exist to deal with the impact of comorbidity with regard to CUAs: the additive, minimum and multiplicative approach (J. Hanmer et al., 2009). Given the availability of utility score lists for single conditions in the Medical Expenditure Panel Survey (MEPS), these approaches could be used to generate utilities for the comorbidities that would be of interest when performing cost-utility-analyses (CUAs) (A. Basu et al., 2008; C.N. McIntosh et al., 2010). In addition, besides the three models reported by Hanmer, there are also other models in use, like the maximum and average model.

1.2. Objectives and research questions

This thesis aims to examine the impact of comorbidity or multimorbidity on EQ-5D3L health utility scores in a nationally representative sample of the US. The research question is the following:
''What is the effect of comorbidity on the utility scores that people assign to their health status? And what is the best model to predict the utility scores for comorbidity based on the adjusted R-Squared (R^2) and the Mean Squared Error (MSE)?''

The secondary questions are the following:

- What is the best model to predict the impact of comorbidity on a patient's health-state utility score?
- Do the impact and optimal model differ by type of comorbidity (-ies)?

1.3. Chapter's overview

Chapter 2 'Background' discusses in detail what is already known on (1) health state utility values, (2) comorbidities, and (3) models that are currently being used for the estimation of health state utility values. Moreover, an overview of previous empirical research is provided. Chapter 3 'Methods' describes the study design, the main characteristics of the dataset that was used, the variables that were used (including a conceptual representation), inclusion criteria of respondents, and the data analysis. Chapter 4 'Results' presents a factual reporting of the study results including Tables to summarize the statistical information. Finally, Chapter 5 '(Discussion)' gives an interpretation of the goodness of fit and explains the findings relative to the theoretical framework and literature. Furthermore, the strengths, limitations, validity and reliability of this thesis are discussed including recommendations for future research.

2. Background

2.1 What is utility?

In health economics, ‘utility’ is the measure of the preference or value that an individual or society gives a particular health state. Utility is generally rated a number of 0 (representing death or the worst imaginable health state) and 1 (representing perfect health) (National Institute for Health and Care Excellence, n.d.). EQ-5D is one of the instruments used to measure the HR-QOL and utility values of respondents. The EQ-5D provides a simple descriptive profile and a single index value for each health status (Rabin R. et al., 2001). The EQ-5D self-reported questionnaire includes a visual analog scale (VAS), which records the respondent’s perceived and self-rated health status on a graded (0-100) scale, with higher scores representing better HR-QOL. It also includes the EQ-5D descriptive system, which comprises five dimensions of health: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression (Rabin R. et al., 2001). The level of severity reported on each of the EQ-5D dimensions together determine a unique health state. Health states are converted into a weighted health state index by applying EQ-5D preference weights (‘tariffs’) elicited from general population samples. These weights lie on a scale on which full health has a value of 1 and dead (or the worst imaginable health state) a value of 0 (J. W. Shaw & J. A. Johnson, 2005; S. Grandy & Fox K. M., 2008).

2.1.1. What is comorbidity?

Comorbidity is the presence of one or more health conditions in addition to the primary condition occurring in the same person at the same time (N. Hoeymans et al., 2001; C.N. McIntosh et al., 2010). It can be due to a simple co-occurrence of diseases or result from a pathophysiological relationship between coexisting conditions (A. M. Broemeling et al., 2005; D. A. Feeny et al., 2005; M. Fortin et al., 2004; R. Gijsen et al., 2001). This study will focus on unrelated comorbidity. Unrelated comorbidities are those in which health conditions occur simultaneously by chance alone (e.g., prostate. cancer and diabetes mellitus), whereas related comorbidities – which is the more common type – consists of health conditions that are systematically associated (G. Andrews, 1998). An example of dependent comorbidity is where multiple health conditions share a common underlying risk factor, as is the case with obesity. Sometimes conditions arise as complications of others, as in the situation where a person with diabetes mellitus develops diabetic retinopathy. Dependent comorbidities can also be concurrent side effects of treatment, for example, the combination of impotence and urinary incontinence that often follows radical prostatectomy (T. C. Kao et al., 2000; C. N. McIntosh et al., 2010).

2.1.2. Approaches to deriving utility scores for comorbidities

Mathematical models: additive, multiplicative, minimum, maximum and average

There are several models to estimate the impact of comorbidity (-ies) in decision analysis. One method used to incorporate comorbidity into utility measurement is the mathematical function which combines the utility scores for each of the individual disease conditions involved. Given the availability of utility score lists for single conditions, this approach could be used to generate utilities for the majority of the comorbidities that would be of interest when performing CUAs. The three mathematical models for combining utility scores are: (1) the additive, (2) multiplicative, and (3) minimum models (C.N. McIntosh et al., 2010).

The *additive model* can be expressed as follows: $U_{(CM)} = 1.0 – \{(1–U_1) + (1– U_n)\}$. $U_{(CM)}$ represents the overall utility score for the comorbidity of i single health conditions (i = 1 through N), and U_i is the utility score for the $_i$th condition implicated in the comorbidity. This approach assumes that the separate HR-QOL losses resulting from each of the individual conditions involved in the comorbidity are additive. For example, for two single conditions having utility scores of 0.90 and 0.80, the overall utility score for the corresponding comorbid state under the additive model is: 1.0 – [(1.0 – 0.90) +

(1.0 – 0.80)] = 0.70. Each additional health condition acquired lowers the overall HR-QOL by a fixed amount irrespective on initial HR-QOL (C.N. McIntosh, 2010).

The *multiplicative model* combines utility scores in the following manner: $U_{(CM)} = \{U_1 * \ldots * U_n\}$. This approach assumes that a comorbid disease increases the patient's utility loss. For example, those in full health who developed a condition having a utility score of 0.90 would maintain 90% of their original level of HR-QOL. If they developed a second condition for which the utility score was 0.80, the new utility score for the comorbid combination would be 0.72 (0.80 * 0.90), representing 72% of full health (C.N. McIntosh, 2010).

The *minimum model* can be expressed as follows: $U_{(CM)} = \min\{U_1, \ldots U_n\}$. It assumes that comorbidity has no additional detrimental effect on the utility values of individuals with an existing health condition (R. Ara & J.E. Brazier, 2010). In this model an individual with multiple conditions is modelled by recognizing only the health condition with the minimum single condition utility score. A disutility is applied that can vary depending on the baseline utility modelled (R. Ara & J.E. Brazier, 2010). For example, the impact of diabetes on health utility could be –0.4 and the impact of congestive heart failure could be –0.10. The impact on health utility in an individual with both health conditions would be calculated by taking the impact that gives the minimum score (-0.40) (C.N. McIntosh, 2010).

Other models are also possible, such as the *maximum* model with the following formula: $U_{(CM)} = \max\{U_1, \ldots U_n\}$, and the *average* model with the following formula $U_{(CM)} = \text{mean}\{U_1, \ldots U_n\}$. The maximum model takes the health condition with the maximum single condition health utility score and the average model takes the average score of the health utility scores of the single conditions.

Purification baseline

Currently, there is no consistency in the baseline used when estimating health utility values for chronic health conditions. The purified baseline is obtained by dividing all health utility values by the mean health utility values obtained from individuals with none of the conditions in a dataset. Using a baseline of perfect health overestimates, the utility associated with health conditions. Therefore, results generated from analyses using a baseline of perfect health are not comparable to those generated using an adjusted baseline. The ideal baseline would be the health utility values associated with not having a particular health condition (R. Ara et al., 2010).

2.2. Previous empirical research

The Panel on Cost-Effectiveness in Health and Medicine (PCEHM) and the Committee to Evaluate Measures of Health Benefits for Environmental, Health, and Safety Regulation (CEMHB) recommends the use of generic HRQOL instruments with preference-based scoring systems to quantify health (M.E. Suarez-Almazor et al., 2000). When performing CUA, an estimate of the impact of delaying or removing a given health condition on health state utility values is required. Instead, researchers often use health state utilities from previously published data rather than performing primary data analysis (R. Ara et al., 2011). Evidence base in this area is relatively small and there is no consensus on the most appropriate model. Moreover, comparisons of findings reported in the literature can be difficult because of differences in study designs such as the preference-based measure estimated, differences in datasets and populations and the statistics used to compare the estimated values (R. Ara et al., 2010).

Table 1 summarizes published studies on the estimation of the effect of comorbidity on health state utility values and comparing the methods for predicting the estimated health state utility values. In a study from 2010 by Hanmer, the additive, minimum and multiplicative models were used to model the SF-6D health utility scores for health states involving comorbidity using data from the Medicare Health Outcomes Survey (HOS). In this study 15 chronic conditions were included. The multiplicative model was the best performing model. There are other models tested, such as the direct utility elicitation, indirect utility measurement, hybrid mathematical model, maximum limit model, adjusted decrement estimator, linear regression model and the multiple regression analysis. In the majority of the studies, the multiplicative model was chosen as the most popular and best performing model compared to the other models. The additive model and the minimum model also performed well in some studies. Moreover, previous studies have also shown that comorbidity has an effect on the health state utility values. For instance, as the number of comorbidities increased, the health state utility values decreased. However, previous studies did not test the maximum and average models.

Table 1. *Previous empirical research*

PAPER AUTHOR(S) / YEAR	UTILITY	COMORBIDITY	EFFECT COMORBIDITY ON UTILITY	TYPE OF MODEL
J. HANMER ET AL., 2010	The models were tested using the SF-6D health utility score of the Medicare HOS dataset.	15 conditions were included (e.g., vision impairment etc.)	Most of the health conditions had an impact of –0.02 or –0.03 on SF-6D health utility scale. For the best performing models, combinations of 2, 3 or 4 conditions only showed an average bias of –0.004 in predicted mean score.	Additive, minimum and multiplicative models were used. Additive model performed well for combinations of 7 or less health conditions. The best performing model in the study is the multiplicative model.
P.W. SULLIVAN ET AL., 2012	EQ-5D health utility scores were used from the MEPS.	The total number of chronic conditions for each individual was calculated based on Clinical Classification Categories codes.	Chronic conditions have a significant deleterious impact on EQ-5D index scores.	Multiple regression analysis was used. Multiplicative model performed best.
C.N. MCINTOSH ET AL., 2010	Article provides key facts about utility.	Article provides key facts about comorbidity.	Comorbidity must be accounted for in utility measurement to avoid biased estimates.	Direct utility elicitation, indirect utility measurement, additive, multiplicative, minimum model, hybrid mathematical model. The multiplicative model is the most popular one and performed the best.
J.A. HAAGSMA ET AL., 2011	Data from the EQ-5D instrument for injury patients with and without comorbidity was used.	Six persisting diseases that were most often reported and comorbid injury were selected.	When the number of comorbid diseases increased, the EQ-5D disability weights of injury patients also increased significantly. Heart disease had the lowest mean EQ-5D disability weight and osteoarthritis had the highest.	Maximum limit, additive and multiplicative approach were tested. The goodness-of-fit of available comorbidity adjustment approaches was high. However, the maximum limit approach seemed to fit less well than the other two models.

R. ARA & J. BRAZIER ET AL., 2011	Health state utility values were obtained from EQ-5D and SF-6D data	32 subgroups with comorbid health conditions were identified.	All scores from the subgroups with comorbid health conditions were smaller than those from the subgroups with single health conditions.	The additive, multiplicative and minimum methods, the adjusted decrement estimator, and a linear regression model were used. The linear model had the most accurate scores for comorbid health conditions. The additive model underestimated, and the minimum method overestimated the actual SF-6D scores respectively.
W. FLANAGAN ET AL., 2006	A utility-based measure of HR-QOL was used.	Study looked at estimates of health utility for the 20 most prevalent comorbid conditions.	Persons with 2 chronic conditions had an average age- and sex-standardized health utility scores ranging from –0.01 (SE=0.00) to 1.00 (SE=0.00).	Findings support the use of multiplicative model.
H. TÜZÜN ET AL., 2015	Level of QOL concerning health was determined by WHOQOL-BREF.	DM and hypertension, musculoskeletal diseases with the accompanying diseases, and cardiovascular diseases with the accompanying diseases.	The QOL decreased with the increasing number of up to 3 chronic diseases in physical domain and with an increasing number of up to 2 diseases in other domains. DM-hypertension comorbidity had a strong negative effect on QOL.	9 different linear regression models were constituted.
R. ARA ET AL., 2012	EQ-5D data from the Health Survey for England were used to compare the health utility values.	39 individual chronic clinical conditions and 15 grouped chronic clinical conditions were used.	A small proportion (6,2%) of the mean health utility values for cohorts with a combined health condition were greater than one of the mean health utility values for the corresponding single health conditions.	The additive, multiplicative and minimum methods, the adjusted decrement estimated and a linear regression were used. The additive and minimum methods performed very poorly in data of study. Study recommends using the multiplicative model together with univariate sensitivity analyses.

Review of literature (Table 1) indicates that the effect of comorbidity or multimorbidity on the health utility values is still unknown. The multiplicative, additive and minimum models are commonly used approaches for the estimation of health state utility values for comorbidities. There is a lack of consensus on the most appropriate method (R. Ara et al., 2012).

3. Research methods

3.1. Study design

The present study is a cross-sectional registry-based study. The impact of comorbidity burden on EQ-5D index scores was examined by estimating the additive, multiplicative, minimum, maximum and average model on existing data in the MEPS database. MEPS Household Component (HC) is a nationally representative survey of the US civilian non-institutionalized population with oversampling of Black and Hispanic respondents. MEPS collects detailed information on demographic characteristics, health conditions and health conditions preference-based HR-QOL via the EQ-5D3L questionnaire. The sample design of the MEPS HC survey includes stratification; clustering; multiple stages of selection; and disproportionate sampling. MEPS sampling weights incorporate adjustment for the complex sample design and reflect survey non-response and population totals from the Current US Population Survey. Medical condition diagnoses in MEPS are based on the International Classification of Disease, Ninth Revision (ICD-9) and Tenth Revision (ICD-10) codes (P. W. Sullivan et al.,2016; Agency for Healthcare Research and Quality, n.d.). In this study the 2002 MEPS data that included the health status survey was pooled. Health status survey questions, including the EQ-5D3L, are administered by paper and pencil. For individuals who are unable to respond, the questionnaire is completed by a proxy. The proxy is generally a family member and head of household. However, specific data on the relationship of the proxy to the respondent is not available.

The 2002 pooled MEPS datafile consists of 39,165 individuals with valid EQ-5D preference weights, of which 27,283 were aged≥18 (69.7%). For the analysis, only adults were selected. To ensure that the condition was experienced while the EQ-5D was administered and because of the assumption that chronic diseases have a plausible effect on utilities, only chronic conditions were included in the analysis of the MEPS database. People who have no chronic conditions (48.9%) were excluded from the analysis.

In this study seven persisting chronic diseases that were most often reported were selected: diabetes, asthma, arthritis, high blood pressure, stroke, joint pain and emphysema. Cases were excluded if they had 'inapplicable/unknown/refused' recorded for education (0.6%), income (0%) and chronic condition (4%).

3.2. MEPS variables

- Age in MEPS represents the exact age as of 12/31/02, calculated from date of birth. For the analysis, age in years was recoded into the following subgroups:
  - 18-26 (=1)
  - 26-35 (=2)
  - 36-45 (=3)
  - 46-55 (=4)
  - 56-65 (=5)
  - 66-75 (=6)
  - 76+ (=7)

- Sex in MEPS was categorized into male (=1) and female (=2).

- Income in MEPS was constructed by dividing family income by the applicable poverty line (based on family size and composition), with percentages grouped into 5 categories: poor/negative (=1), near poor (=2), low income (=3), middle income (=4), and high income (=5).

- Educational level in MEPS indicates the highest degree of education attained and was categorized into eight groups: no degree (=1), GED (=2), high school diploma (=3), bachelor's degree (=4), master's degree (=5), doctorate degree (=6), and other degree (=7). GED stands for General Education Diploma and is equivalent to a high school diploma.

- Chronic conditions were recoded into the following dummy variables: yes (=1); as in the presence of the chronic condition, and no (=0); as in the absence of the chronic condition.

3.2.1. Other variables

Aside from the variables that were included in the 2002 MEPS datafile, other variables were made for the statistical analysis:

- SES encompasses income and education and was classified into low SES (=1), middle SES (=2) and high SES (=3). The data exploration showed that income and educational level were moderately correled (r=0.384). Thus, to avoid possible collinearity and confounding in the regression analysis, educational and income were combined into one variable SES. Educational level was recoded into four groups: low educated (=1), middle educated (=2), high educated (=3), and no degree (=4). Low educated included the group 'no degree', middle educated combined the groups 'GED' and 'high school diploma'. Furthermore, high educational level combined the groups 'bachelor's degree, master's degree and doctorate degree'. Subsequently, family income was recoded into three groups: low income (=1), middle income (=2) and high income (3). Low income combined the groups 'poor/negative' and 'near poor'. People with a low income and low educational achievement were considered as having a low SES. People with a low income and middle educational achievement were also considered as having a low SES. People with a low income and a high educational achievement were considered as having a middle SES, and vice versa.

- Number of chronic conditions (NCC) represents the number of chronic conditions that a person is diagnosed with. This variable was created by summing the seven chronic conditions (diabetes, arthritis, asthma, emphysema, stroke, joint pain and high blood pressure). The variable NCC consists of 8 levels, ranging from 0 to 7. For instance, if a person has one chronic condition NCC = 1.

- EQ5D utility was generated according to the D1 model (Table 1) developed by N. Luo et al., 2007, which is a preference-based scoring function for the EQ-5D3L based on time trade-off (TTO) valuations from the general adult US population. The D1 model consists of 10 dummies (2 for each dimension), 3 ordinal variables representing (squared) numbers of dimensions that are in level 2 or 3, and an ordinal variable called 'D1'. The D1 term represents the number of dimensions beyond the first that are not in level 1 (value range, 0-4) (N. Luo et al., 2007).'

Table 1 Specifications of the D1 and N3 Models

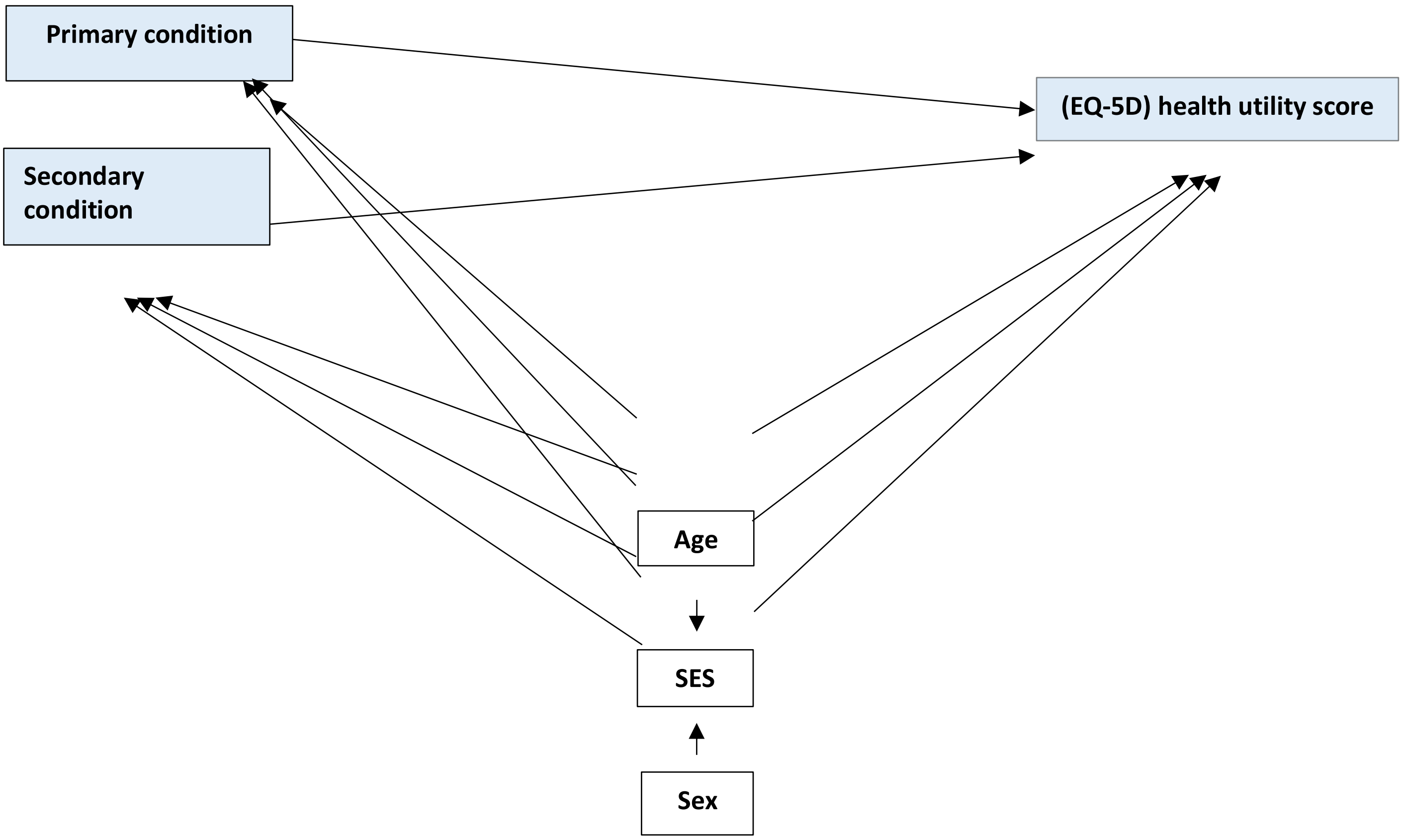
Parameter	Definition	Estimate	
		D1 Model[10]	N3 Model[5]
Constant		NA	0.081
M2	1 if mobility is level 2; 0 otherwise	0.146	0.069
M3	1 if mobility is level 3; 0 otherwise	0.558	0.314
S2	1 if self-care is level 2; 0 otherwise	0.175	0.104
S3	1 if self-care is level 3; 0 otherwise	0.471	0.214
U2	1 if usual activities is level 2; 0 otherwise	0.140	0.036
U3	1 if usual activities is level 3; 0 otherwise	0.374	0.094
P2	1 if pain/discomfort is level 2; 0 otherwise	0.173	0.123
P3	1 if pain/discomfort is level 3; 0 otherwise	0.537	0.386
A2	1 if anxiety/depression is level 2; 0 otherwise	0.156	0.071
A3	1 if anxiety/depression is level 3; 0 otherwise	0.450	0.236
N3	1 if any dimension is level 3; 0 otherwise	NA	0.269
D1	Number of dimensions beyond the first not in level 1	–0.140	NA
I2-squared	Squared number of dimensions beyond the first in level 2	0.011	NA
I3	Number of dimensions beyond the first in level 3	–0.122	NA
I3-squared	Squared I3	–0.015	NA

Note: The dependent variable of both models was disutility (i.e., 1 – preference value). NA, not applicable.

Table 2. *Parameters and estimates of D1 model (N. Luo et al., 2007)*

3.3. Directed Acyclic Graph

***Figure 2**: DAG*



In this analysis, the independent variables (*or exposures*) are primary condition and secondary conditions (comorbidities), and the dependent variable (*or outcome*) is (EQ-5D) health utility score. The adjustment variables are age, sex and SES. The DAG in Figure 2 shows the relationship between the exposures, outcome and adjustment variables, which takes the form of lines going from one variable to another. These lines are directed, which means to say that they have a single arrowhead indicating their effect (Barrett, 2010). For instance, primary condition and secondary condition can have an influence the health utility score. This effect is influenced (possibly confounded) by age, sex and SES. In addition, age and sex can have an effect on SES. This DAG was used to identify confounding and sources of bias, which is particularly important for this thesis. If confounding is present, it can cause an over- or underestimation of the observed association between the exposure and outcome (K. Alexander et al., 2013).

3.3.1. Association between confounders and exposure and outcome in the literature

Impact age, sex and SES on comorbidity

The DAG shows that comorbidity is influenced by age, sex and SES. The association that age and SES have with comorbidity is further explained in a study from 2015 by Tüzün; as the world's population gets older and the frequency of chronic diseases increases, the frequency of comorbidity will also increase (H. Tüzün et al., 2015). In addition, comorbidity is more frequently observed in low socioeconomic groups. Furthermore, in another study from 2008 by Wanfu, it is explained that men and women have differences in health status, which could lead to differences in comorbidities. (W. Wanfu, 2008).

Impact age, sex and SES on utility

The DAG also shows that utility is influenced by age, sex and SES. This is confirmed in a study from 2019 by Cubi-Molla et al in which is stated that older respondents value health profiles lower than younger age groups. This could lead to differences in the valuation of health across groups of people of different ages (P. Cubi-Molla et al., 2019). Furthermore, they found that women have a lower QOL which could lead to differences in health utility scores. In addition, studies have indicated that people in lower socioeconomic groups have a lower QOL than people in higher socioeconomic groups. The QOL got better when either the income or the SES increased (H. Tüzün et al., 2015; M. Keyvanara et al., 2015). This could also result in lower or higher health utility scores.

Impact sex on SES

The impact of gender on SES is explained in a study by the American Psychological Association in which is stated that SES affects the overall well-being and quality of life for women (American Psychological association, 2010). In the US, for instance, men are paid more than women, despite similar levels of education and equivalent fields of occupation. Reduced income for women, coupled with longer life expectancy and increased responsibility to raise children, increase the probabilities that women will face economic disadvantages.

3.4. Correlations between income and educational level

Table 3 presents the relationship between the variables 'family income as percent of poverty line (*income*)' and 'highest educational level achieved (*education*)' in the MEPS datafile. The bivariate correlation analysis showed that income and education are moderately correlated with a Spearman correlation of 0.40 (P<0.01). To avoid confounding in the multiple regression analysis, income and education were combined into one variable: socio-economic class (SES). People in the groups 'poor/negative, near poor and low income' were combined into one group: low income group. People with a low income and low educational achievement were considered as having a low SES; people with a low income and middle educational achievement were considered as having a middle SES; people with a low income and high educational achievement were considered as having a high SES, and vice versa. The variable SES was included in the OLS regression models. Furthermore, the correlations between sex, SES and age were assessed. The bivariate analysis showed no correlation between these 3 variables.

Subsequently, reference groups were made for SES and age. The new variables that were used in the unadjusted OLS regressions were the following:

***Table 3**. Relationship income and education for patients with NCC ≥ 2*

	Family income as percent of poverty line			
Highest degree attained	Low income	Middle income	High income	Total
Low educated	4267 (15.7%)	2009 (7.4%)	771 (2.8%)	7047 (26%)
Middle educated	4419 (16.3%)	4689 (17.3%)	4139 (15.3%)	13247 (48.8%)
High educated	547 (2.0%)	1209 (4.5%)	3311 (12.2%)	5067 (18.7%)
Other degree	334 (1.2%)	621 (2.3%)	803 (3.0%)	1758 (6.5%)
Total	9567 (35.3%)	8528 (31.4%)	9024 (33.3%)	27119 (100%)

3.4.1. Correlations between Sex, SES and Age

For the unadjusted OLS regressions, reference groups were made for the variables SES and age. Therefore, SES was divided in two groups (SESM and SESL) and age was divided into 6 age groups (AgeGroup1-6). To assess the correlations between these variables, a Bivariate analysis was conducted. The Spearman values indicate that there is (almost) no correlation.

Table 4. *Correlations between Sex and subgroups of SES and Age.*

			SES2L	SES2M	AgeGroup1	AgeGroup2	AgeGroup3	AgeGroup4	AgeGroup5	AgeGroup6	SEX
Spearman's rho	SES2L	Correlation Coefficient	1.000	-.774**	.129**	.014*	-.057**	-.109**	-.052**	.023**	.061**
		Sig. (2-tailed)	.	.000	.000	.024	.000	.000	.000	.000	.000
		N	26096	26096	25361	25361	25361	25361	25361	25361	26096
	SES2M	Correlation Coefficient	-.774**	1.000	-.030**	-.026**	.035**	.043**	.020**	-.005	-.040**
		Sig. (2-tailed)	.000	.	.000	.000	.000	.000	.001	.414	.000
		N	26096	26096	25361	25361	25361	25361	25361	25361	26096
	AgeGroup1	Correlation Coefficient	.129**	-.030**	1.000	-.209**	-.220**	-.203**	-.157**	-.127**	-.012*
		Sig. (2-tailed)	.000	.000	.	.000	.000	.000	.000	.000	.047
		N	25361	25361	27283	27283	27283	27283	27283	27283	27283
	AgeGroup2	Correlation Coefficient	.014*	-.026**	-.209**	1.000	-.249**	-.230**	-.178**	-.144**	-.017**
		Sig. (2-tailed)	.024	.000	.000	.	.000	.000	.000	.000	.005
		N	25361	25361	27283	27283	27283	27283	27283	27283	27283
	AgeGroup3	Correlation Coefficient	-.057**	.035**	-.220**	-.249**	1.000	-.242**	-.187**	-.151**	.002
		Sig. (2-tailed)	.000	.000	.000	.000	.	.000	.000	.000	.795
		N	25361	25361	27283	27283	27283	27283	27283	27283	27283
	AgeGroup4	Correlation Coefficient	-.109**	.043**	-.203**	-.230**	-.242**	1.000	-.172**	-.140**	-.011
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.	.000	.000	.058
		N	25361	25361	27283	27283	27283	27283	27283	27283	27283
	AgeGroup5	Correlation Coefficient	-.052**	.020**	-.157**	-.178**	-.187**	-.172**	1.000	-.108**	-.002
		Sig. (2-tailed)	.000	.001	.000	.000	.000	.000	.	.000	.730
		N	25361	25361	27283	27283	27283	27283	27283	27283	27283
	AgeGroup6	Correlation Coefficient	.023**	-.005	-.127**	-.144**	-.151**	-.140**	-.108**	1.000	.017**
		Sig. (2-tailed)	.000	.414	.000	.000	.000	.000	.000	.	.005
		N	25361	25361	27283	27283	27283	27283	27283	27283	27283
	SEX	Correlation Coefficient	.061**	-.040**	-.012*	-.017**	.002	-.011	-.002	.017**	1.000
		Sig. (2-tailed)	.000	.000	.047	.005	.795	.058	.730	.005	.
		N	26096	26096	27283	27283	27283	27283	27283	27283	39165

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

3.5. Data analysis

Firstly, descriptive statistics were used to report data about the MEPS sample's characteristics (age, gender, highest education attained, family income), utilities, (number of) chronic conditions. The number of respondents and corresponding percentage was reported for each variable. Secondly, the mean age, mean EQ-5D score (and standard deviation), median EQ-5D score (and interquartile range) and the number of chronic conditions (NCC) were computed for each variable. For the statistical analyses the mean EQ-5D scores (calculated according to the D1 model) for each subgroup (individuals who have only one primary health condition, i.e. in absence of any comorbidity) was calculated. The mean EQ-5D scores for each single primary chronic condition are presented in Table 5. Table 5 shows that individuals who are only diagnosed with diabetes have a utility score of 0.86.

***Table 5**. Utilities for the single primary health conditions*

Single primary disease	Actual EQ-5D score
Diabetes	0.86
Asthma	0.90
Arthritis	0.84
Emphysema	0.83
Stroke	0.80
High blood pressure	0.88
Joint pain	0.86

Subsequently, the mean EQ-5D scores for the single chronic conditions were used to compute the estimated EQ-5D scores using the additive, multiplicative, minimum, average and maximum models. For the calculation of the estimated EQ-5D scores, a baseline of perfect health (1.0) was used.

For the unadjusted analysis, the following regression equation was used:

*U(PD+COM1+COM2) = b0 + b1*U(PD, COM1, COM2) + e*

For the adjusted analysis, the following regression equation was used:

*U(PD+COM1+COM2) = b0 + b1*U(PD, COM1, COM2) + b4*COM1 + b5*COM2 + b6* COM3 + b7 * COM4 + b8*COM5 + b9*COM6 + b10*COM7 +b11*AGE + b12*AGE + b13*AGE + b14*AGE + b15*AGE + b16*AGE + b17*SEX + b18*SES + b19*SES e*

3.5.1. Goodness of fit

The impact of comorbidity on EQ-5D health state utility values is estimated with multiple Ordinary Least Squares (OLS) regression analysis using the following goodness of fit parameters: the adjusted R2 and MSE. The impact of the variables on utility was expressed as beta-coefficients (95%CI and p-value), controlling for age, sex, SES, and presence of comorbidity (diabetes, asthma, arthritis, joint pain, stroke, high blood pressure and emphysema). The beta-coefficients (95% CI) represents the estimated impact. A total of five separate regression models were made for each of the mathematical models. The choice of the best performing model in estimating the impact of the actual EQ-5D scores on comorbidity was based on the adjusted R2 and MSE. The latter measures the average of the square of errors. The higher the MSE, the larger the error. Error in this case means the difference between the EQ-5D health state utility values as record in the MEPS database according to the D1 model and predicted EQ-5D health state utility values based on the specification of the regression model. In addition, the adjusted R2 measures how close the estimated EQ-5D health state utility values (fitted regression line) match the observed EQ-5D health state utility values. The adjusted R2 value (ranging between 0 to 1) gives the measure of how much variance is explained by the model. A high value can indicate that the model fits the data. However, the adjusted R2 can't determine whether the model is biased (J. Frost, 2019). Hence, residual plots for each model had to be plotted. For the analysis of the data, the Statistical Package for the Societal Sciences (SPSS) version 26.0 was used. For further details on the statistical approach see Appendix 1.

4. Results

4.1. Descriptive statistics

Patient characteristics

There was a total of 39,165 individuals in the pooled 2002 MEPS data of which 27,283 were adults (aged≥18) (69.7%). People aged between 36 and 45 years were the biggest group and people older than 76 were the smallest group (Table 6). About 30% of the sample had high school as the highest educational achievement. In the descriptive statistics, there are several relationships that appear. Increasing comorbidity burden was associated with older age. Moreover, females tend to be slightly older, with greater comorbidity. In addition, there is an association between education and age as people with a higher education tend to be older than people with a lower education. The most prevalent reported condition is joint pain with more than 33% of the sample reporting this condition. The least reported condition is emphysema with a prevalence of 1.6% in the sample. More than half of the respondents reported not having any comorbidities (NCC=0 and 1), while 14.4% reported having one comorbidity (NCC=2). As the number of chronic conditions increased, less people reported having multiple chronic conditions. Furthermore, people with a higher family income had lower comorbidity levels.

Table 6. *Demographic characteristics mean age and mean NCC of respondents in the 2002 Medical Expenditure Panel Survey*

Variables	N (%)	Mean age (years)	Mean NCC
MEPS (all adults)			
Age groups (years) as of 12-31-2002		45	
18-25	4250 (15.6%)	21	0.26
26-35	5234 (19.2%)	31	0.41
36-45	5663 (20.8%)	41	0.68
46-55	4973 (18.2%)	50	1.18
56-65	3206 (11.8%)	60	1.69
66-75	2195 (8.0%)	70	2.03
76+	1762 (6.5%)	81	2.20
Sex			
Male	12568 (46.1%)	32	0.86
Female	14715 (53.9%)	35	1.08
Educational level			
No degree	7047 (19.5%)	45	1.04
GED	1324 (3.4%)	44	1.13
High school diploma	11923 (30.7%)	44	0.98
Bachelor's degree	3424 ((8.8%)	45	0.83
Master's degree	1312 (3.4%)	49	0.95
Doctorate degree	331 (0.9%)	51	0.86
Other degree	1758 (4.5%)	45	1.00
Family income as percent of poverty line			
Poor/negative	6936 (17.7%)	28	1.13

Near poor	2225 (5.7%)	31	1.09
Low income	6654 (17.0%)	32	1.01
Middle income	12123 (31.0%)	34	0.93
High income	11227 (28.7%)	39	0.93
Chronic conditions			
Diabetes	1993 (7.3%)	60	3.02
Arthritis	5506 (20.2%)	60	2.76
Asthma	2487 (9.1%)	45	2.39
High blood pressure	6530 (23.9%)	59	2.44
Stroke	721 (2.6%)	67	3.46
Joint pain	9085 (33.3%)	53	2.23
Emphysema	427 (1.6%)	65	3.67
Number of chronic conditions (NCC)			
1	6433 (23.6%)	45	1.00
2	3941 (14.4%)	54	2.00
3	2231 (8.2%)	61	3.00
4	997 (3.7%)	64	4.00
5	275 (1.0%)	64	5.00
6	56 (0.2%)	65	6.00
7	6 (0.0%)	69	7.00

*NCC = number of chronic conditions

Figure 3. *Distribution of the EQ-5D index score computed according to D1 model.*

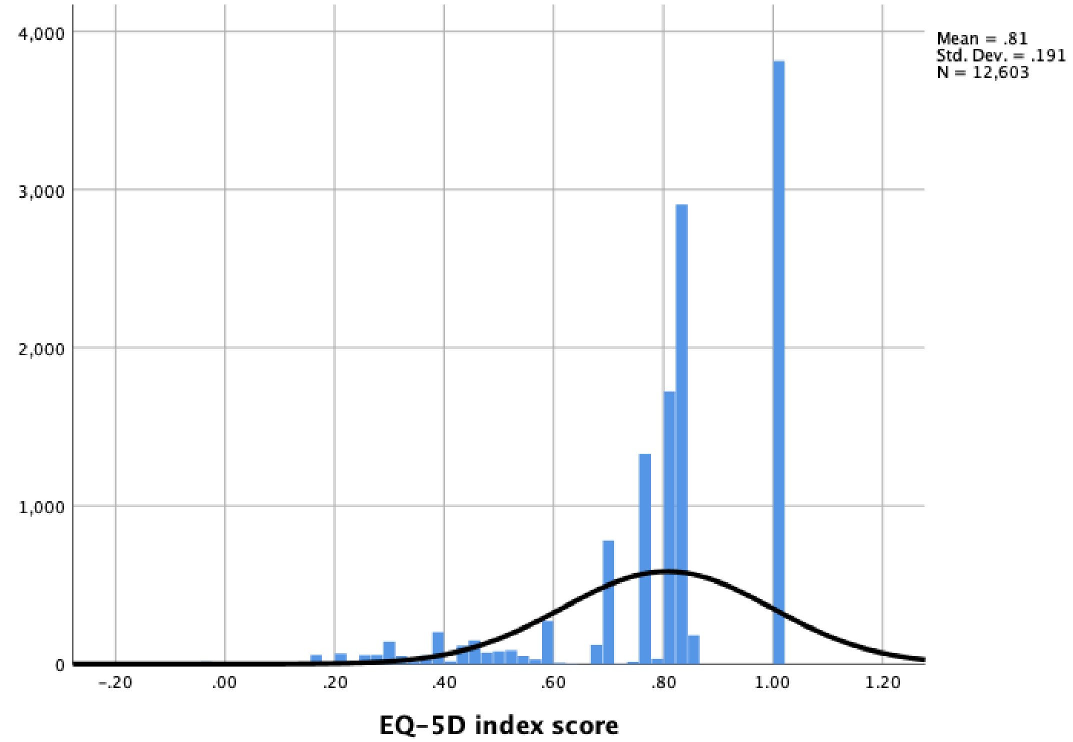


Figure 3 shows the normal distribution curve of the EQ-5D utility score computed according to the D1 model of Nan Luo (2007). The distribution curve is slightly skewed to the right. In addition, the data points cluster more towards the right side of the scale (score point 1).

Table 7. *Relationship between mean/median EQ-5D scores and age groups, sex, income, education, and (number of) chronic conditions excluding NCC = 0*

Variable	Mean EQ-5D score index (SD)	Median EQ-5D score index
Age groups (years)		
18-25	0.92 (0.125)	1.00
26-35	0.91 (0.140)	1.00
36-45	0.88 (0.161)	1.00
46-55	0.84 (0.181)	0.83
56-65	0.82 (0.188)	0.83
66-75	0.80 (0.178)	0.83
76+	0.74 (0.210)	0.79
Sex		
Male	0.88 (0.167)	1.00
Female	0.85 (0.177)	0.84
Family income as percent of poverty line		
Poor/negative	0.79 (0.220)	0.83
Near poor	0.82 (0.206)	0.83
Low income	0.85 (0.186)	0.84
Middle income	0.87 (0.163)	0.84
High income	0.90 (0.134)	1.00
Education		
No degree	0.83 (0.204)	0.83
Graduate equivalency degree (GED)	0.83 (0.193)	0.83
High school diploma	0.86 (0.166)	0.84
Bachelor's degree	0.90 (0.131)	1.00
Master's degree	0.91 (0.128)	1.00
Doctorate degree	0.91 (0.142)	1.00
Other degree	0.88 (0.158)	0.86
Chronic conditions		
Diabetes	0.74 (0.217)	0.84
Arthritis	0.74 (0.205)	0.80
Asthma	0.80 (0.204)	0.83
High blood pressure	0.78 (0.196)	0.82
Stroke	0.68 (0.229)	0.77
Joint pain	0.78 (0.198)	0.82
Emphysema	0.67 (0.224)	0.71
Number of chronic conditions		
1	0.87 (0.155)	0.84
2	0.80 (0.182)	0.83

3	0.73 (0.202)	0.78
4	0.67 (0.204)	0.76
5	0.61 (0.219)	0.71
6	0.54 (0.221)	0.59
7	0.61 (0.192)	0.71

Table 7 presents the mean and median EQ-5D health utility scores for the sociodemographic factors including NCC and chronic conditions of patients with a chronic condition. Firstly, it is notable that older patients tend to have lower EQ-5D scores. Moreover, females had a mean EQ-5D score of 0.79 versus an EQ-5D score of 0.83 for men. Increasing family income as percent of poverty line and higher levels of education were associated with greater EQ-5D scores. In addition, emphysema and stroke had the lowest EQ-5D scores of 0.67 and 0.68, respectively, whereas asthma had the highest EQ-5D score of 0.80. Furthermore, the EQ-5D score increased with the number of comorbid diseases. It is also notable that people with 7 chronic conditions have a higher EQ-5D utility score than people with 6 chronic conditions.

Table 8. *Comparison of the accuracy of EQ-5D health state utilities using additive, multiplicative, minimum, maximum and average model.*

Using a baseline of perfect health	**Actual**	**Additive**	**Multiplicative**	**Minimum**	**Maximum**	**Average**
Mean EQ-5D score	0.72	0.53	0.65	0.85	0.88	0.86
Minimum EQ-5D score	–0.10	–0.03	0.33	0.80	0.86	0.83
Maximum EQ-5D score	1.00	0.76	0.79	0.86	0.90	0.88
Range	1.10	0.79	0.47	0.06	0.04	0.05
Mean error	0.01	0.004	0.003	0.0005	0.0003	0.0002

4.1.1. Scatterplots

Spread estimated values

Figures 3A to 3E present the scatterplots for the EQ-5D health utility scores calculated according to the D1 model (X-axis) and the estimated EQ-5D health utility scores calculated using the five models (Y-axis). The diagonal line shows the combination of utilities where estimated and actual utility scores are exactly equal. The more accurately the model predicts, the closer the dots are to the diagonal, the lower the MSE and the higher the adjusted) R2 (D. B. Figueiredo Filho et al., 2019).

The estimated utility range using the multiplicative model (0.33-0.79) is smaller than the actual utility range (-0.10–1.00). For the additive model this is (-0.03-0.76), which is more similar to the range of the actual utilities. The estimated utilities for the maximum, average and minimum model are (0.86-0.90) (0.86-0.88), (0.80-0.86) respectively. The maximum, minimum and average model overestimated the actual utilities. As can be shown in Figures 3C, 2D and 2E, the estimated utilities are spread above the diagonal. For the multiplicative and additive models, the estimated utilities are spread around the diagonal. However, both models underestimated higher utility values and overestimated lower utility values. In the maximum, minimum and average model all utilities are overestimated. It is noticable that a large group of individuals had an actual health utility score of around 0.80. However, the additive model and the (a little bit less in) multiplicative model weren't able to predict these higher values accurately and underestimated the actual values.

***Figure 3A**. Scatterplot Additive model*

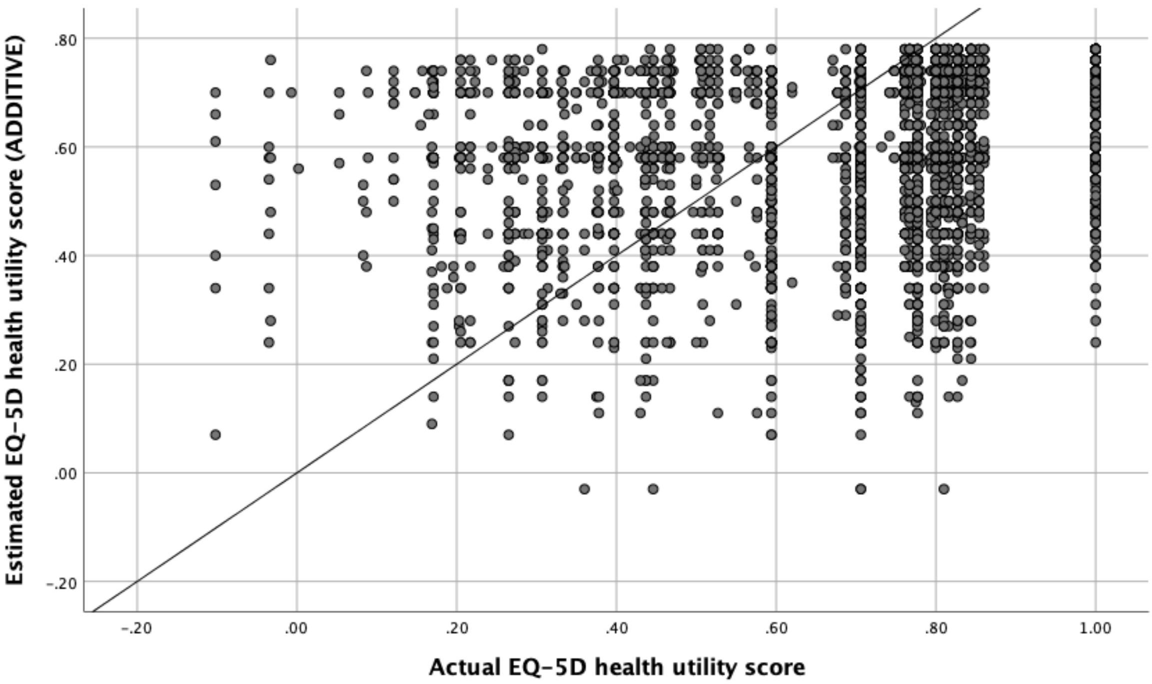


Figure 3B. Scatterplot Multiplicative model

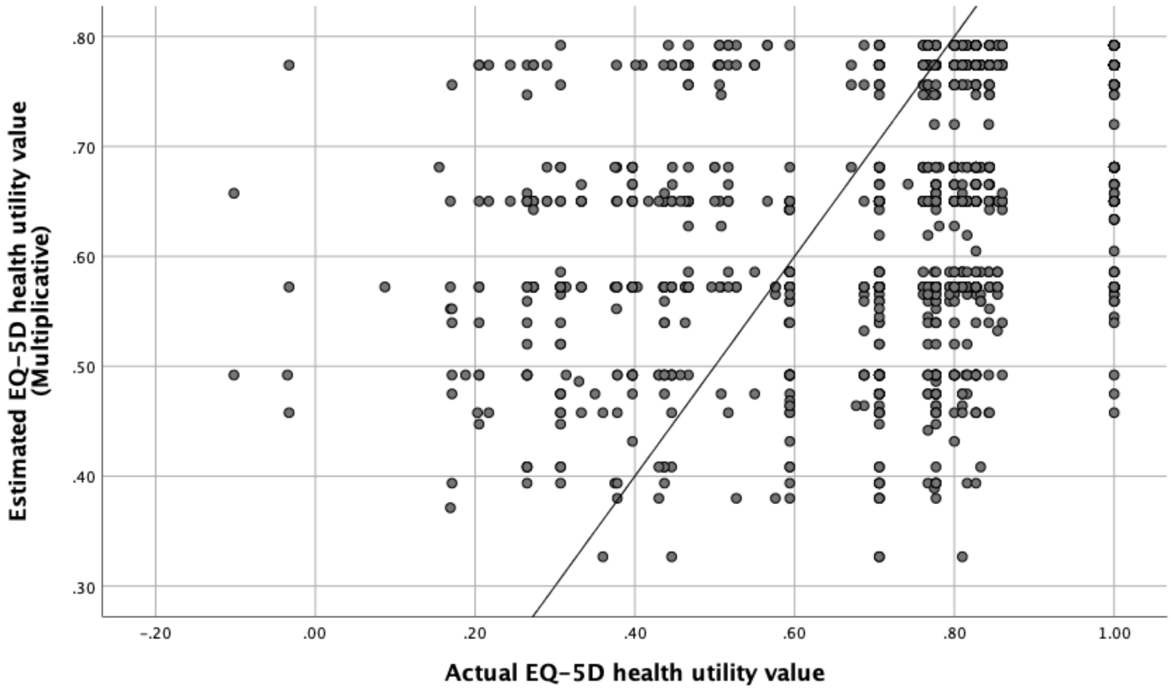


Figure 3C. Scatterplot Minimum model

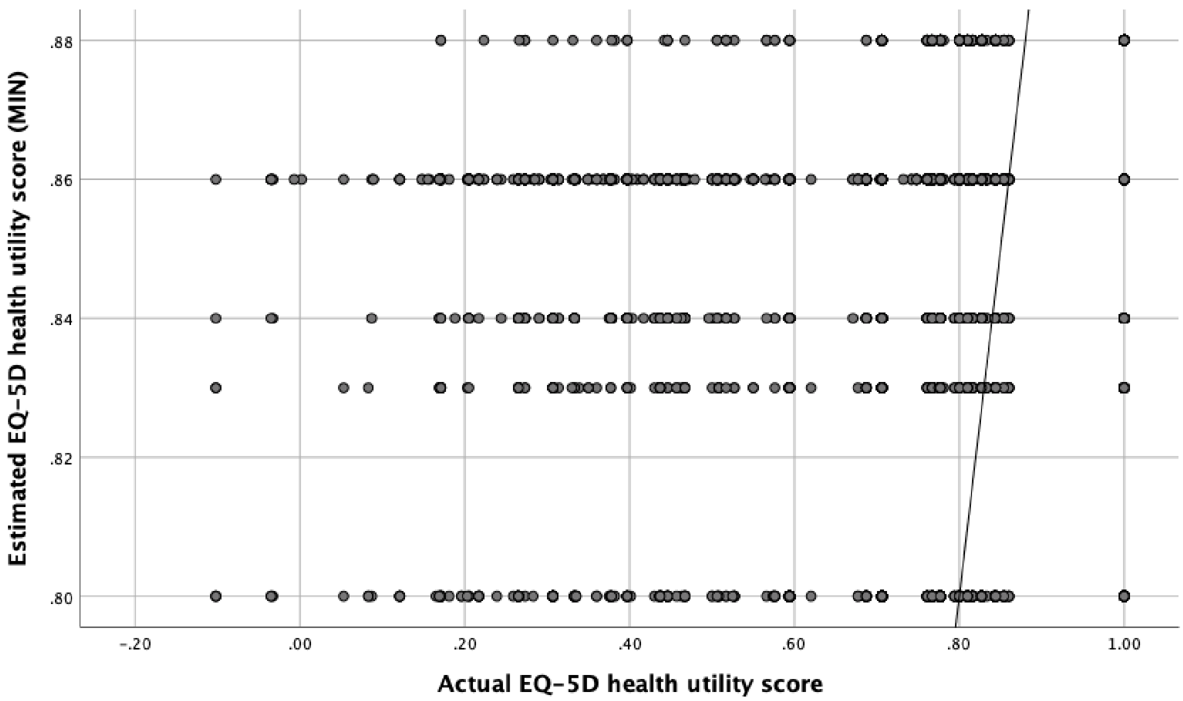


Figure 3D. Scatterplot Maximum model

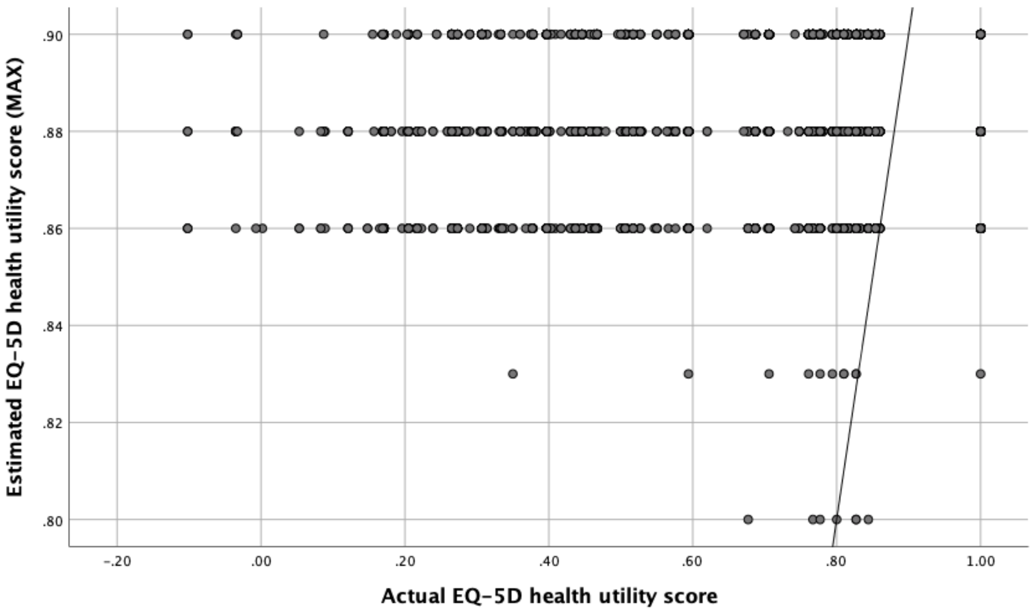
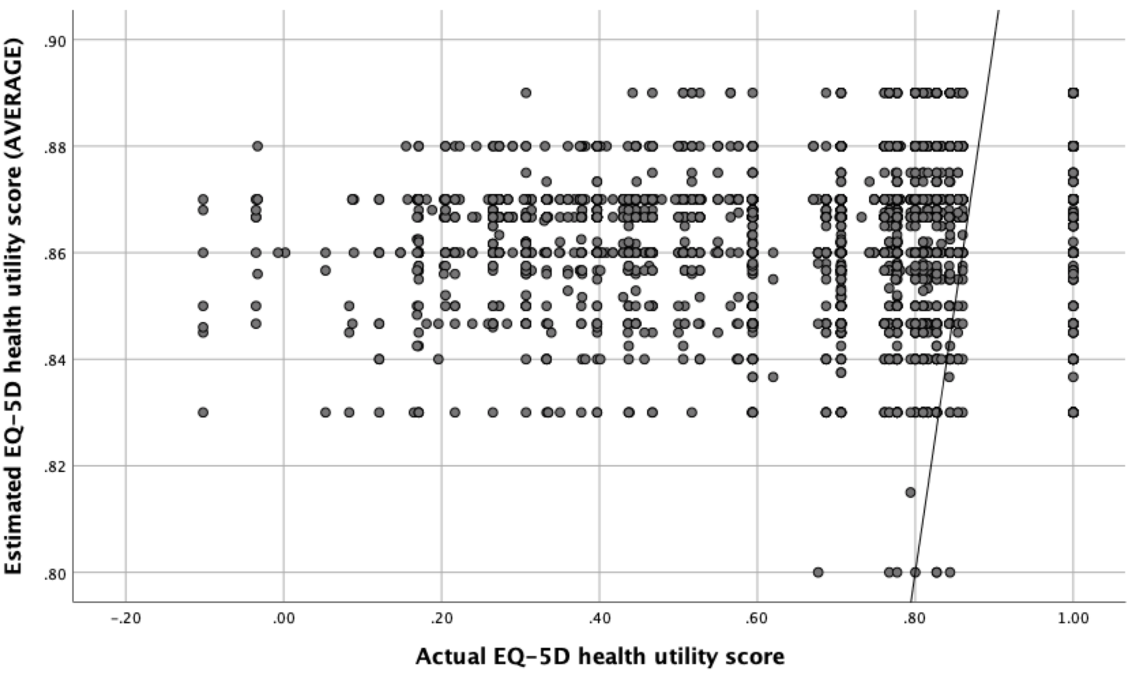


Figure 3E. Scatterplot Average model



4.2. Linear regression models

4.2.1. Goodness of fit

Table 10 includes the mean and 95% CI for the independent variables and the goodness of fit parameters – adjusted R2 and MSE – associated with each model adjusted for age, sex, SES and comorbidities. The parameters in this Table were used to evaluate which of the unadjusted and adjusted models fits best using a comorbid pair of all seven diseases. For the unadjusted analysis, the multiplicative model produced the largest adjusted R2 (0.131) compared to the other four models. Conversely, the maximum model produced the lowest adjusted R2 (0.00). The additive model achieved the second highest adjusted R2 (0.094). Moreover, the multiplicative model also produced the lowest MSE (0.035) which indicates that the multiplicative model estimates a better fit of the mean EQ-5D health utility value (referred to as 'actual' EQ-5D-score from here on) compared to the other models using comorbid pair of seven diseases. However, the additive and minimum models also had similar overall MSE between the actual and estimated EQ-5D utility scores of 0.037 and 0.039 respectively. The maximum model produced the largest MSE in the unadjusted analysis (0.041), which means that it performed worse in predicting the actual EQ-5D health utility values compared to the other models.

When adjusting for age, sex, SES and chronic conditions, the multiplicative model had the highest adjusted R2 (0.198) and the smallest MSE (0.033). The other models had an adjusted R2 of 0.162 and 163. The additive, minimum and average models had the same overall MSE values (0.034).
Furthermore, the maximum method produced again the largest MSE (0.041), which indicates that it also didn't perform well in predicting the actual EQ-5D score in the adjusted analyses for all seven comorbid pairs. Although the maximum model with adjustments and without adjustments, had the same MSE (0.041), one can conclude that the models overall performed better in the adjusted analyses than in the unadjusted analysis. The multiplicative model achieved the smallest MSE (0.033). This means that the multiplicative model performed better than the other models in predicting the EQ-5D health utility values accurately (D.B. Figueiredo Filho et al., 2019)

Independent variable	Additive		Multiplicative		Minimum		Maximum		Average	
For all diseases	95% CI	ß	95% CI	ß	95% CI	ß	95% CI	ß	95% CI	ß
Constant	0.424, 0.556		-0.435, 0.937		-0.441, 0.762		-0.172, 1.581		-1.315, 0.640	
Age (years)										
18-25	0.023, 0.092	0.041*	-0.012, 0.097	0.041	0.020, 0.090	0.040	0.022, 0.092	0.040*	0.019, 0.640	0.041*
26-35	-0.007, 0.040	0.018*	-0.012, 0.097	0.019	-0.008, 0.039	0.040	-0.007, 0.040	-0.019	-0.009, 0.039	0.019
36-45	-0.026, 0.010	0.012	-0.019, 0.035	0.019	-0.026, 0.010	0.019	-0.026, 0.010	-0.004	-0.026, 0.010	0.019
46-55	-0.018, 0.013	0.009	-0.023, 0.028	-0.013	-0.017, 0.014	-0.013	-0.017, 0.013	0.010	-0.017, 0.014	-0.013
56-65	-0.010, 0.020	0.008	-0.007, 0.043	-0.004	-0.010, 0.020	-0.004	-0.010, 0.020	0.055*	-0.009, 0.021	-0.004
66-75	0.013, 0.044	0.008	-0.006, 0.058	0.010	0.013, 0.044	0.010	0.013, 0.044	-0.006	0.013, 0.044	0.010
Sex	-0.012, 0.007	0.008	-0.019, 0.009	0.055	-0.012, 0.007	0.055*	-0.012, 0.007	-0.123	-0.012, 0.007	0.055*
SESM	-0.066, -0.035	0.005	-0.066, -0.021	-0.006	-0.065, -0.034	-0.006	-0.066, -0.035	-0.329	-0.065, -0.034	-0.006
SESL	-0.150, -0.188	0.008*	-0.155, -0.109	-0.123	-0.149, -0.117	-0.123*	-0.150, -0.118	-0.089	-0.149, -0.117	-0.123*
Diabetes	-0.054, -0.032	0.008*	-0.032, 0.107	-0.329	-0.149, -0.117	-0.329	-0.054, -0.032	-0.188	-0.051, -0.027	-0.329*
Asthma	0.007, -0.033	0.007*	-0.022, 0.077	0.060	-0.036, -0.009	-0.089*	-0.072, --0.010	-0.081	-0.061, -0.029	-0.089*
Arthritis	xx	xx	-0.093, 0.077	0.041*	-0.091, -0.069	-0.081*	-0.093, -0.068	-1.88	-0.090, -0.068	-0.062*
Joint pain	-0.026, 0.008	0.009	-0.067, 0.075	-0.17	-0.086, -0.059	-0.188*	-0.100, -0.058	-0.152	-0.088, -0.9161	-0.190*
Stroke	0.002, 0.044	0.011	-0.076, 0.128	0.032*	-0.066, 0.020	-0.152*	-0.096, -0063	-0.111	-0.074, -0.009	-0.152*
Emphysema	-0.016, 0.031	0.012*	-0.062, 0.107	0.009	-0.083, -0.030	-0.111*	-0.100, -0.058	-0.088	-0.085, -0.035	-0.111*
High blood pressure	0.014, 0.037	0.006*	-0.038, 0.080	0.061*	-0.047, -0.026	-0.098*	-0.060, -0.022	-0.098	-0.063, -0.035	-0.089*
Ucm	0.445, 0.579	0.331*	0.055, 1.539	0.331*	0.277, 1.657	0.024	-0.670, 1.349	0.505	0.418, 2.671	-0.085*
Adjusted R2 for unadjusted analysis	0.094		0.131		0.037		0.000		0.025	
MSE for unadjusted analysis	0.037		0.035		0.039		0.041		0.040	
Adjusted R2 for adjusted analysis	0.162		0.198		0.163		0.162		0.163	

MSE for adjusted analysis	0.034	0.033	0.034	0.034	0.034

***P (<0.01)**
Ucm = estimated EQ-5D health utility variables computed according to each of the models
-- = variable removed from regression because of arthritis patients with no registered utilities
Age reference = 76+
Sex reference = male
SESM = middle SES, SESL = low SES
SES reference = SESH (high SES)
Diabetes, asthma, arthritis, joint pain, stroke, emphysema, high blood pressure reference show the impact when each of these conditions is present (reference: not present)

Table 10. *Goodness of fit of actual comorbid utility values and comorbid utility values predicted with the five mathematical models adjusted for age, sex, SES and type of comorbidity*

Table 11 (summarizing table for 'separate diseases') gives an overview of the goodness of fit measures when each of the 7 diseases is analysed separately. When adjusted for sex, age, SES and chronic conditions, all diseases show a higher adjusted R2 except for 'diabetes'. The latter shows a higher adjusted R2 without the adjustments. Moreover, for the multiplicative model, emphysema has a higher adjusted R2 without the adjustments.

Disease	**Results Adjusted R2 and MSE for (un)adjusted analysis**	**Additive model**	**Multiplicative model**	**Minimum model**	**Maximum model**	**Average model**
Diabetes	Adjusted R2 for unadjusted analysis	0.185	0.211	0.186	0.184	0.186
	MSE for unadjusted analysis	0.038	0.036	0.038	0.038	0.038
	Adjusted R2 for adjusted analysis	0.150	0.175	0.044	0.005	0.035
	MSE for adjusted analysis	0.045	0.037	0.045	0.047	0.045
Asthma	Adjusted R2 for unadjusted analysis	0.159	0.163	0.104	--	0.118
	MSE for unadjusted analysis	0.038	0.038	0.041	--	0.040
	Adjusted R2 for adjusted analysis	0.251	0.250	0.250	--	0.246
	MSE for adjusted analysis	0.035	0.035	0.035	--	0.035
Arthritis	Adjusted R2 for unadjusted analysis	0.071	0.077	0.038	0.017	0.008
	MSE for unadjusted analysis	0.039	0.044	0.040	--	0.040
	Adjusted R2 for adjusted analysis	0.138	0.158	0.139	0.138	0.138
	MSE for adjusted analysis	0.036	0.041	0.036	0.036	0.036
Joint pain	Adjusted R2 for unadjusted analysis	0.086	0.135	0.039	0.004	0.018
	MSE for unadjusted analysis	0.038	0.036	0.040	0.041	0.038
	Adjusted R2 for adjusted analysis	0.157	0.209	0.158	0.157	0.158

	MSE for adjusted analysis	0.035	0.033	0.035	0.035	0.035
Stroke	adjusted R2 for unadjusted analysis	0.068	0.056	--	0.002	0.008
	MSE for unadjusted analysis	0.049	0.054	--	0.053	0.052
	Adjusted R2 for adjusted analysis	0.098	0.082	0.098	0.098	0.096
	MSE for adjusted analysis	0.047	0.052	0.047	0.048	0.048
Emphysema	adjusted R2 for unadjusted analysis	0.083	0.091	0.009	0.004	-0.003
	MSE for unadjusted analysis	0.045	0.044	0.049	0.049	0.050
	Adjusted R2 for adjusted analysis	0.190	0.162	0.190	0.188	0.189
	MSE for adjusted analysis	0.040	0.040	0.040	0.040	0.040
High blood pressure	Adjusted R2 for unadjusted analysis	0.127	0.147	0.045	0.004	0.040
	MSE for unadjusted analysis	0.035	0.033	0.039	0.040	0.039
	Adjusted R2 for adjusted analysis	0.183	0.205	0.185	0.183	0.184
	MSE for adjusted analysis	0.033	0.032	0.033	0.033	0.034

-- = variable removed from regression because of patients with no registered utilities.

***Table 11**. Goodness of fit (adjusted-R2, MSE) of actual comorbid utility values and comorbid utility values predicted with the five models adjusted for age, sex, SES and type of comorbidity (summarizing table for 'separate diseases')*

5. Discussion

The purpose of this thesis was to assess the performance of five models in predicting the EQ-5D health utility values for comorbidities. Health state utility scores have been developed primarily for single health conditions. Therefore, inaccuracies in health measurements are likely to occur when comorbidity is ignored in the estimation of the health utility scores. Currently, there is no consensus on the most appropriate method. Hence, the need for this specific research in this area is evident. Although the adjusted R2 was very small, the multiplicative showed better results than the other models.

''What is the effect of comorbidity on the utility scores that people assign to their health status?
The presence of comorbidity has a significant impact on the EQ-5D health utility values. The beta-coefficients indicate that if a person has diabetes, asthma, arthritis, joint pain, stroke, emphysema or high blood pressure, the EQ-5D health utility score is likely to decrease. This effect was also shown when multiple comorbidities were selected. This is shown by the negative values of the beta-coefficients.

''What is the best model to predict the utility scores for comorbidity based on the Adjusted R2 and the MSE?''
When conducting an unadjusted analysis for all seven diseases, the multiplicative model produced the highest adjusted R2 (0.131) and the lowest MSE (0.035). Conversely, the maximum model had the lowest adjusted R2 (0.000), indicating that the model explains none of the variability of the response data around its mean. For the adjusted analysis the multiplicative model also had the highest adjusted R2 (0.198) and the lowest MSE (0.033). This indicates that the multiplicative model performed better than the other models in predicting the actual EQ-5D health utility values for comorbidity based on the adjusted R2 and MSE. However, the estimated values for adjusted R2 are very low.

''Do the impact and optimal model differ by type of comorbidity (-ies)?
The multiplicative and additive models gave the most accurate results for the separate diseases (Table 11). This means that the optimal model doesn't differ by the type of comorbidity. In all separate diseases, the adjusted R2 was higher in the adjusted analysis (with the highest R2 (0.250) in the multiplicative and additive (0.251) models), except for the diabetes group. The latter performed better in the unadjusted analysis.

5.1. Interpretation findings relative to theoretical framework

The goodness of fit parameters in this thesis didn't perform well in estimating the actual EQ-5D health state utility values. In this thesis the adjusted analysis resulted in a higher adjusted R2. The reason is that – when adding more independent variables into the regression – the adjusted R2 starts to increase. When selecting more independent variables (for example, race) into the regression, the adjusted R2 for the multiplicative model increased from 0.190 to 0.199. However, the adjusted R2 is still low. Using other parameters or other models might give better results.

5.1.2. Interpretation findings relative to literature

It is difficult to compare the findings of this thesis with the findings of other studies. The main reason is because other studies used other parameters (e.g., standard errors, mean absolute difference) and other utility measurements. However, the finding that the multiplicative and additive models are the best performing models is in accordance with a study from 2010 by Hanmer et al (2010). In this study the additive, minimum and multiplicative methods were used to model the SF-6D health utilities for comorbidities. The authors indicated that the multiplicative model demonstrated the best performance in reconstructing the observed utilities for the comorbidities and that it had the best

MSE. The finding that the models showed bad performance in predicting the actual EQ-5D health state utility value is in accordance with a study from 2008 by Fu et al. In this study the minimum model showed the best performance (ß=1.029) while the additive model showed the worst performance (ß=0.084). However, the authors stated that – in general – all models showed bad performance and that the minimum model is '*the best of a bad lot'*. Furthermore, the authors stated that it was difficult to determine why this was the case in their study. This can be due to the differences in the utility measurement and the health conditions that were included in their analysis. This was also the case in this thesis: although the multiplicative had the highest adjusted R2, the models didn't perform well in predicting the utility value. These discrepancies could stem from the method effects.

5.2. Strengths and limitations

Strengths

In this thesis real population data was used from a large representative study with heterogeneous characteristics. The MEPS dataset included: different age groups (0-85 years), different social classes (low, middle, high SES), and multiple/different chronic diseases and comorbidities. Therefore, it accurately reflects the characteristics of the population and it can provide more accurate mean values J. Young, 2021). Moreover, both unadjusted and adjusted regressions were performed. The linear regressions showed that the goodness of fit parameters increased when adjusting for age, sex, SES and comorbidities. This shows that adjustments are very important as it can take into account any factors that can influence the outcome.

Limitations

MEPS does not include information on the severity of a condition, whilst it has been demonstrated that adjusting for disease severity improves the associations between morbidity burden and HRQOL outcomes (MEPS, 2004). Furthermore, some of the conditions that are categorized as chronic, may not have a chronic impact for all respondents reporting the condition. Additionally, for individuals who cannot complete the EQ-5D, MEPS allows the use of proxy respondents. Approximately 12% of all MEPS responses to the EQ-5D were completed by proxies These proxy respondents may rate health differently from the individuals themselves and may underestimate the HR-QOL (Andresen et al., 2001). This may introduce a source of bias. Moreover, MEPS is based on self-report and the ability of survey respondents to report accurate condition data may be a source of bias for the conditions included. This bias may be exacerbated in Blacks and Hispanics, who may differ in reporting of levels of illness and disability. These groups are more likely to report worse general physical health (C. Brach et al., 2006). There is also evidence that self-reported conditions may be underreported in general population (MEPS, n.d.)

Another issue is that individuals from the general population in the MEPS database reported more than one chronic condition, but still reported an EQ-5D index score of 1 (full health). An explanation could be that these individuals are in a low-severity disease stage or are effectively treated and thus have low(er) score. This phenomenon was also seen in the scatterplots of the models. Individuals who reported an actual index score of 1, had an estimated index score of lower than 1. This indicates that the models underestimate higher values of EQ-5D health utility scores.

5.3. Validity and reliability

Selection bias
For arthritis patients (n=1694), only (n=835) had registered EQ5D utilities. This could be due to the severity of their disease. For example, patients with arthritis could have difficulty writing. As a result, patients who are 'very sick' are not included in the data which might result selection bias. Moreover, when conducting regression models adjusted for age, sex, SES and chronic diseases, the independent variable 'arthritis' was removed from the analysis and therefore wasn't included in the regression model. This might also lead to bias in the estimates.

Information bias
In this analysis, the D1 model of Nan Luo was used in predicting the tariffs. In 2011, a newer version of the 5L (EQ-5D-3L (3L)) was developed. The main difference between the standard 3L and the 5L is that it has five dimensions now instead of three (B. Janssen et al., 2018) The question is whether this implies that the results of the analysis are outdated or not. Studies have compared the descriptive systems of the 3L and the 5L in terms of their measurement properties, ceiling effects and evenness, and reliability and validity (M.F. Janssen et al., 2013; C.B. Agborsangaya et al., 2014). The findings indicate that the 5L descriptive system has more precise and better measurement properties compared with the 3L. Moreover, the 3L overestimated health problems and underestimated utilities. This indicates that the D1 model is a valid tariff as it leads to less bias. Therefore, the results are not outdated.

Number of participants
The number of participants in the 2002 pooled MEPS datafile consisted of 39,165 (100%) individuals with valid EQ-5D preference weights. However, this number became much smaller. When only selecting individuals who were 18 years or older the number of participants became 27,283 (69.7%)/ When selecting individuals who also have 2 chronic conditions or more, the number of participants became 7506 (19.7%). Thus, only 19.7% of the pooled datafile was included in the statistical analysis.

Coding
In the analysis several variables had to be combined, made or recoded. The correlation and internal consistency got smaller. This could have an affect the reliability of the data quality.

5.4. Recommendations
Based on the findings of the statistical analysis, the EQ-5D health utility score decreases in case of comorbidities. This requires attention from healthcare professionals and policymakers. For instance, when patients experience several (chronic) diseases, integrated care should be encouraged among healthcare professionals to optimize healthcare delivery (I. Heide et al., 2015), as studies have shown that integrated care can increase the utility of patients with comorbidity (Van Duin et al., 2017). Accordingly, stimulating integrated care for patients with comorbidities would therefore be important.

The multiplicative, additive, minimum, maximum and average models didn't show good performance in estimating the actual utility. To verify the impact of comorbidity on the health utility score, other models (*for instance the hybrid mathematical model*) should be used alongside the traditional models. Studies have shown that the hybrid mathematical model showed a good fit to the actual utilities for comorbidities (Basu et al., 2008). This model appears promising, however further research is recommended to verify and specify these findings.

References

Agency for Healthcare Research and Quality. (n.d.). Medical Expenditure Panel Survey Home. Retrieved February 19, 2021, from https://www.meps.ahrq.gov/mepsweb/

American Psychological Association. (2010). *Fact Sheet: Women & Socioeconomic Status*. Retrieved February 5, 2021, from https://www.apa.org/pi/ses/resources/publications/women

Andrews, G. (1998). Burden of disease. *British Journal of Psychiatry*, *173*(2), 123–131. https://doi.org/10.1192/bjp.173.2.123

Ara, R., & Brazier, J. E. (2010). Populating an Economic Model with Health State Utility Values: Moving toward Better Practice. *Value in Health*, *13*(5), 509–518. https://doi.org/10.1111/j.1524-4733.2010.00700.x

Ara, R., & Brazier, J. (2011). Estimating Health State Utility Values for Comorbid Health Conditions Using SF-6D Data. *Value in Health*, *14*(5), 740–745. https://doi.org/10.1016/j.jval.2010.12.011

Ara, R., & Brazier, J. (2012). Comparing EQ-5D Scores for Comorbid Health Conditions Estimated Using 5 Different Methods. *Medical Care*, *50*(5), 452–459. https://doi.org/10.1097/mlr.0b013e318234a04a

Barra, M., & Augestad, L. A. (2015). Examining the relationship between health-related quality of life and increasing numbers of diagnoses. *Quality of Life Research*, *24*(12), 2823–2832. https://doi.org/10.1007/s11136-015-1026-3

Bell, C. M., & Chapman, R. H. (2001). An Off-the-Shelf Help List. *Medical Decision Making*, *21*(4), 288–294. https://doi.org/10.1177/0272989x0102100404

Bo Hu, & Fu, A. Z. (2010). Predicting Utility for Joint Health States: A General Framework and a New Nonparametric Estimator. *Medical Decision Making*, *30*(5), E29–E39. https://doi.org/10.1177/0272989x10374508

Cubi-Molla, P. (2018). A note on the relationship between age and health-related quality of life assessment. *Quality of Life Research*, *28*(5), 1201–1205. https://doi.org/10.1007/s11136-018-2071-5

Dale, W., & Basu, A. (2007). Predicting Utility Ratings for Joint Health States from Single Health States in Prostate Cancer: Empirical Testing of 3 Alternative Theories. *Medical Decision Making*, *28*(1), 102–112. https://doi.org/10.1177/0272989x07309639

Figueiredo Filho, D. B. (2011). What is R2 all about? *Leviathan (São Paulo)*, (3), 60. https://doi.org/10.11606/issn.2237-4485.lev.2011.132282

Flanagan, W., & McIntosh, C. N. (2006). Deriving utility scores for co-morbid conditions: a test of the multiplicative model for combining individual condition scores. *Population Health Metrics*, *4*(1), 1–8. https://doi.org/10.1186/1478-7954-4-13

Fu, A. Z., & Kattan, M. W. (2008). Utilities Should Not Be Multiplied. *Medical Care*, *46*(9), 984–990. https://doi.org/10.1097/mlr.0b013e3181791a9c

Gijsen, R., & Hoeymans, N. (2001). Causes and consequences of comorbidity. *Journal of Clinical Epidemiology*, *54*(7), 661–674. https://doi.org/10.1016/s0895-4356(00)00363-2

Grandy, S., & Fox, K. M. (2008). EQ-5D visual analog scale and utility index values in individuals with diabetes and at risk for diabetes: Findings from the Study to Help Improve Early evaluation and management of risk factors Leading to Diabetes (SHIELD). *Health and Quality of Life Outcomes*, *6*(1), 18. https://doi.org/10.1186/1477-7525-6-18

Haagsma, J. A., & van Beeck, E. F. (2011). The effect of comorbidity on health-related quality of life for injury patients in the first year following injury: comparison of three comorbidity adjustment approaches. *Population Health Metrics*, *9*(1), 1–8. https://doi.org/10.1186/1478-7954-9-10

Hanmer, J. (2010). Three methods tested to model SF-6D health utilities for health states involving comorbidity/co-occurring conditions. *Journal of Clinical Epidemiology*, *63*(3), 331–341. https://doi.org/10.1016/j.jclinepi.2009.06.013

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence Based Nursing*, *18*(3), 66–67. https://doi.org/10.1136/eb-2015-102129

Josten, J. (2017). The David Foster Wallace Reader by David Foster Wallace. *Middle West Review*, *3*(2), 87–89. https://doi.org/10.1353/mwr.2017.0009

Kao, T.C., & Cruess, D. F. (2000). Multicenter patient self-reporting questionnaire on impotence, incontinence and stricture after radical prostatectomy. *Journal of Urology*, *163*(3), 858–864. https://doi.org/10.1016/s0022-5347(05)67819-6

McIntosh, C. N. (2010). Utility Scores for Comorbid Conditions: Methodological Issues and Advances. *Handbook of Disease Burdens and Quality of Life Measures*, 359–380. https://doi.org/10.1007/978-0-387-78665-0_20

Medical Expenditure Panel Survey. (2004). MEPS HC-070: 2002 Full Year Consolidated Data File. Retrieved July 17, from https://meps.ahrq.gov/data_stats/download_data/pufs/h70/h70doc.pdf

National Institute for Health and Care Excellence. (n.d.). Glossary. Retrieved February 14, 2021, from https://www.nice.org.uk/Glossary?letter=U#Utility

Suarez-Almazor, M. E., Kendall, C., Johnson, J. A., Skeith, K., & Vincent, D. (2000). Use of health status measures in patients with low back pain in clinical settings. Comparison of specific, generic and preference-based instruments. *Rheumatology*, *39*(7), 783–790. https://doi.org/10.1093/rheumatology/39.7.783

Shaw, J. W., & Johnson, J. A. (2005). US Valuation of the EQ-5D Health States. *Medical Care*, *43*(3), 203–220. https://doi.org/10.1097/00005650-200503000-00003

Sullivan, P. W., & Ghushchyan, V. H. (2016). EQ-5D Scores for Diabetes-Related Comorbidities. *Value in Health*, *19*(8), 1002–1008. https://doi.org/10.1016/j.jval.2016.05.018

Sullivan, P. W., & Ghushchyan, V. H. (2012). The Impact of Co-Morbidity Burden on Preference-Based Health-Related Quality of Life in the United States. *PharmacoEconomics*, *30*(5), 431–442. https://doi.org/10.2165/11586840-000000000-00000

Tüzün, H. & Aycan S. (2015). How Quality of Life Changes with Comorbidity and Socioeconomic State for Patients with Chronic Diseases who Apply to Primary Health Care Centres. *Central European Journal of Public Health*, *23*(3), 188–194. https://doi.org/10.21101/cejph.a3990

Wanfu, W. (2008). *Age, SES, and health : old topic, new perspective (a longitudinal perspective about the relationship between SES and health)*. Retrieved February 18, 2021, from University of Texas Libraries: https://repositories.lib.utexas.edu/handle/2152/18316

Van der Heide, I. (2015). *Innovating care for people with multiple chronic conditions in Europe*. Retrieved from https://www.nivel.nl/sites/default/files/bestanden/Rapport-State-of-the-Art-ICARE4EU.pdf

Appendix 1: Statistical approach

1.
PD represents primary disease, and COM1 and COM2 represent two (unrelated) comorbidities that may or not be associated with the primary disease.
The usual approach to estimate the impact of comorbidity on health state utilities is to set up the following regression model. Assuming linearity,
(A.1) U(PD+COM1+COM2) = b0 + b1*U(PD) + b2*COM1 + b3*COM2 + e

with U(PD) as the utility of the primary disease without comorbidity, and COM1, COM2 represent the presence/absence of comorbidity COM1 and COM2 (dummy 0/1 variables). U(PD+COM1+COM2), U(PD), COM1 and COM2 are available in the MEPS dataset.

Impact of COM1 and COM2 on utility (PD+COM1+COM2) can be evaluated from the estimated coefficients of b2 and b3. Goodness of fit is evaluated as adjusted R2 and the root mean squared error (RMSE).

Note the following:
1) The impact of COM1 on U(PD1+…) ≠ impact of COM1 on U(PD2+…)
2) When studying different patient populations, i.e. different PD with or without different COM1, COM2,…

Regarding the latter, U(PD+COM1+COM2) and U(PD) might refer to different patient populations (despite the same PD), in terms of age, sex and perhaps other. So for a fair comparison, (A.1) should probably be adjusted for age, sex and other parameters. In this thesis, the focus on three patient characteristics that are known to impact utility: age, sex and socio-economic status.
Take these effects into account, (A.1) may be changed into

(A.2) U(PD+COM1+COM2) = b0 + b1*U(PD) + b2*COM1 + b3*COM2 + b4*AGE + b5*SEX +
+ b6*SES + e

The impact of COM1 (estimated coefficient b2) and COM2 (estimated coefficient b3) might change when AGE, SEX and SES are included in (A.2).
(A.2) can be used to estimate coefficients b2 and b3 when all these determinants are unrelated, i.e. confounding is absent. However, one cannot be certain in advance that confounding is absent. As Figure 2 (DAG) shows, AGE, SEX and SES *may* be confounding variable and *may* be interrelated, so the assumption of independency underlying (A.2) may not be true.
In that case, conventional linear regression (OLS, ordinary least squares regression) may not be the correct way to estimate coefficients b2 and b3 from (A.2) but other techniques have to be used, e.g. stratified OLS, or IVAR.

2.
Another approach to evaluate the impact of comorbidity on utility is to model the impact using (one of) the models outlined in the thesis, e.g. the additive or multiplicative model. In that case,

(A.3) U(PD+COM1+COM2) = b0 + b1*U(PD) + b2*Ucm(COM1, COM2) + e
with Ucm defined as the impact of comorbidities COM1 and COM2 in presence of primary disease PD, using the selected model.

In analogy to (A.2), (A.4) may read

(A.4) U(PD+COM1+COM2) = b0 + b1*U(PD) + b2*Ucm(COM1, COM2) + b4*AGE + b5*SEX + + b6*SES + e

As in (A.2), confounding may play a role, and other regression models then OLS (e.g. stratified regression analysis, IVAR) may have to be used to estimate coefficients b1 and b2 validly.

The goodness of fit parameters adjusted R2 and RMSE will be used to evaluate which of the unadjusted models (A.1, A.3 type) and adjusted models (A.2, A.4 type) fits best. Note that the answer to that question may depend on the PD and comorbidity characteristics.