The University Of Sheffield.

Bayesian Design of Discrete Choice Experiments for Valuing Health State Utilities

by

Fatimah Ali Aloef

Thesis submitted to the University of Sheffield

for the degree of Doctoral of Philosophy

School of Mathematics and Statistics

January 2015

To my parents and my little family.

Acknowledgments

First I mostly thank my PhD supervisors Prof. Jeremy Oakley and Dr. Eleanor Stillman for all their help and support. I really appreciate their valuable advices during my PhD that contributed to make this thesis possible. It is of my honor to be one of their students and benefit from their experiences.

I wish to express my sincere thanks to Dr. Kevin Walters and Prof. Peter Goos for reading my thesis and providing me with significant and constructive suggestions in the oral examination. This thesis could not have been completed without their generous and professional assistances.

I am also grateful to Prof. John Brazier at the ScHARR of the university of Sheffield for providing us with the data sets for asthma case study and making available for us to use their data. I would like also to thank the staff on the Mathematics department for being friendly and helpful all the time.

I would like also thank all my friends (too many to present here but you know who you are) for being supportive, and providing me the friendship I needed during my studies and understanding my fluctuated mood during the PhD. I am really thankful to my friends Sujunya Boonpradit and Madina Hassan for supporting and encouraging me all the time.

I specially thank my dad, mum, and my brothers and sisters for their support and prayers and for their entire love. Special thanks for my little family, my husband who support me all the time and has faith in me and my intellect even when I felt down. Also my kids Ali and the little angel Sarah where there is no words can describe how much I do love them.

Abstract

This thesis aims to develop an efficient methodology to construct efficient discrete choice experiments (DCEs) for health state utility estimation within the QALY framework.

The use of the QALY measure in health economic evaluation together with methods related to measuring the QALY weight/health state utilities are reviewed in order to establish the fundamental knowledge needed for valuing health. DCEs are used to value health state utilities, which is simpler than other direct valuation methods. Nevertheless, DCEs are still undergoing research to improve their uses in valuing utilities, in particular in designing experiments which are used to construct the DCEs

The main issues with the current choice designs together with design considerations for valuing utilities are identified in this thesis. Advanced work for constructing choice designs, particularly Bayesian optimal design, is reviewed to construct more efficient designs for valuing utilities. Since constructing Bayesian optimal designs requires a prior distribution for the unknown choice model parameters, Bayesian analysis is performed for a real data to obtain appropriate prior distributions.

Constructing Bayesian optimal choice designs for valuing utilities within QALY framework using the existing choice design software is investigated. We find there are limitations because of the design considerations for valuing health state utilities particularly in terms of anchoring utility values into the QALY scale (0-1 scale). We then develop a new algorithm based on modifying the latest advanced choice design algorithms such that they account for the design considerations which overcomes the limitations with the existing design software. Methods for simplifying the choice design questions are also provided.

We demonstrate the use of our design algorithm by constructing Bayesian choice designs for asthma quality of life classification system (AQL-5D), and then investigate the effect of the choice of the prior distribution on the choice of Bayesian designs.

Contents

List of Abbreviations

AQL-5D	Asthma Quality of Life classification with five dimensions/attributes.
CEA	Cost-effectiveness Analysis.
CMA	Cost-minimisation Analysis.
CSPMs	condition specific preference-based measures.
CUA	Cost-utility Analysis.
DALY	disability adjusted life-year.
DCE	discrete choice experiment.
EQ-5D	European qality of life with five dimensions/attributes.
FIM	Fisher information matrix.
GFIM	generalised Fisher information matrix.
GPMs	generic preference-based measures.
HRQoL	health related quality of life.
HTA	Health Technology Appraisal.
HUI	health utility index.
HYE	health years equivalent.
ICER	incremental cost effectiveness ratio.
LBD	level balanced design.
MNL	multinomial logit model.
NICE	National Institute for Health and Care Excellence.
QALY	quality-adjusted life-year.
RUT	random utility theory.
SF-3D	short form with three dimensions.
SG	standard gamble.
TTO	time trade-off.
VCM	variance-covariance matrix.

Chapter 1

Introduction

1.1 Introduction

In this thesis, we tackle the problem of producing efficient surveys (choice experiments) for valuing health outcomes (i.e. health states) to be used within health care evaluation studies. In particular, we develop experimental design strategies to construct efficient choice studies through bringing to bear the latest advanced work in design theory to this important area of research in health economics.

This chapter introduces the main concepts of health economic evaluation, and the use of choice experiments to value health outcomes. In the following section, we present the importance of health economic evaluation in allocating limited health resources, followed by a description for discrete choice experiments in Section 1.3. In Section 1.4, the motivation behind the needs to modify choice design in the health economic field, particularly for valuing health utilities, is discussed briefly. Then, we highlight

the main objective of our research in Section 1.5, and present the outline of the thesis in Section 1.6.

1.2 Economic Evaluation in Health Care

Recently, there has been a rapid advance in modern medicine, where there might be many treatments or health care interventions for a single health condition. There are many conditions to be treated with sometimes scarce health resources (e.g. funds, people, time, facilities, equipment and knowledge). Publicly funded health care organisations, such as the NHS, cannot necessarily offer the best possible treatment for each condition; choices have to be made to allocate limited financial health resources wisely so that the best health benefit is returned.

In making such choices, it is essential to consider both the quality and length of someone's life gained under each treatment. Health decision makers, such as those at the National Institute for Health and Care Excellence (NICE) in England and Wales, take these factors into account when carrying out a Health Technology Appraisal (HTA) on a new health care intervention. The process of the appraisal takes into account the clinical and cost effectiveness of a technology (e.g. a drug) along with other specified considerations through three phases; scope, assessment and appraisal. During the scope phase, the question of interest for a technology appraisal is formulated in order to be addressed later on the assessment procedure. In the assessment process, a health technology is evaluated based on the relevant evidence available to produce an estimate

of a technology's clinical and cost effectiveness for a specific appraisal question and context, while taking into account uncertainty around both quantities. This phase usually consists of two parts; a systematic review and an economic evaluation. The assessment and its analysis are then reported together with additional information supplied by consultees, patient experts and the general public to formulate an appraisal decision. A guidance of the technology appraisal is then produced and sent to the health care providers at the NHS, who revise the recommendations made by NICE and decide whether the new treatment or health intervention provides good value for their money.

A comprehensive analysis is required to ensure that the final guidance issued by the Institute is appropriate and robust. Thus, it is essential to provide a high standard and transparent health economic evaluation study for any new intervention. Health economic evaluation is described by Drummond et al. (2005) as a comparative assessment of alternative courses of action in terms of both their costs and consequences (health care benefits). There are several evaluation techniques to provide evidence for cost effectiveness of an intervention. These techniques, such as Cost-effectiveness Analysis (CEA), Cost-minimisation Analysis (CMA) and Cost-utility Analysis (CUA), mainly differ in the way that health care benefits are measured (see Drummond et al. 2005 for more details about the key distinctions and analysis of each evaluation technique).

CEA has been regarded as the dominant method for decision making policy. In CEA, effectiveness is commonly measured in natural units such as life-year gained or death adverted. The results are expressed as cost per unit of effectiveness. However, using such measures makes it inappropriate to compare interventions with different primary outcomes. For example, kidney transplantation could be compared to heart surgery if the common effect of interest is only life-years gained, but the comparison become difficult if outcome measures differ.

In the 1980s and 1990s, an alternative generic outcome measure that combines the morbidity (quality of life) and mortality (quantity gains) in a single measure was developed. This measure is the quality-adjusted life-year (QALY) which describes

both the quantity and quality of life gained from a particular health intervention. An evaluation method using this measure is termed a cost-utility analysis (CUA). The utility term is used here to refer to individuals' preferences for any particular set of health outcomes. The QALY measure allows for a comparison of the relative effectiveness between interventions for the same disease and interventions from different therapies even if there is no common effect of interest. The results from the CUA are expressed as cost per QALY gained, where one QALY is equivalent to living one year in perfect health.

Though other generic measures have been suggested as an alternative to the QALY, such as health years equivalent (HYE), save young life equivalent and disability adjusted life-year (DALY), few economic evaluations have used these methods as their strengths and weaknesses are not fully established (Drummond et al., 2005). Therefore, many health care decision-maker guidelines recommend the use of analyses that use cost per QALYs gained, as in the reference case of NICE (2008) in UK (NICE guidance to perform HTA), and similar bodies in the USA and Australia (Ryan et al., 2006). The NICE reference case allows the Institution to make a comparison across different health interventions, since it makes sure that the appraisals for all interventions adopt the same approach for the analysis of clinical and cost effectiveness. The reference case specifies the most appropriate methods to conduct an HTA including, for example, whose preference to elicit for valuing health (e.g. patients, carers) and the most appropriate approach to measure health outcomes, together with other aspects of analysis. The methods should be appropriate for the Appraisal Committee's purposes and meet the objective of the NHS of maximising health gain from limited resources. The QALY measure is particulary useful for those organisations where decisions must be made across different interventions which usually have different primary outcomes.

In CUA, making the choice of the most effective intervention depends on comparing the expected change in costs to the expected change in QALYs gained by choosing one intervention over another. This cost effectiveness outcome measure is expressed as

incremental cost effectiveness ratio (ICER):

$$ICER = \frac{\Delta C}{\Delta E} < \kappa, \tag{1.2.1}$$

where ΔC and ΔE are the mean difference in costs and effects, respectively, and κ is the threshold value of willingness to pay per QALY gained (e.g. in NICE reference case (2008) a threshold value of £20,000–£30,000 per QALY gained is used). Therefore, for any two interventions A and B, intervention B is said to be cost effective and provide a good value for NHS budget if $ICER = \frac{C_B - C_A}{E_B - E_A} < \kappa$.

However, a decision made based on the estimated change in cost and effect only would ignore uncertainty associated with these quantities. Thus, sensitivity analysis is often required by the decision maker in order to account for any source of bias and uncertainty around the cost effectiveness model used to inform the estimate of costs and health effects. The model should account for the uncertainty and limitations on the evidence used to estimate costs and effects. NICE (2008) identifies three sources of uncertainty in cost effectiveness analysis: structural uncertainty of the decision model, uncertainty surrounding the sources of collected data, and parameter uncertainty associated with inputs to the model.

It is of interest for health decision makers to minimise bias and uncertainty surrounding the overall decision, to reduce the risk of making inappropriate decisions. This is partly related to uncertainty in the estimated value of cost per QALY gained. Thus, in the following section, we firstly discuss how QALYs are determined and then introduce different methods to value the quality of health outcomes, i.e. the 'Q' part of the QALY, mainly using the discrete choice experiment (DCE) technique.

1.3 Discrete Choice Experiments and Valuing Health

In the NICE reference case for CEA, the value of health effect should be expressed in term of QALYs for the appropriate time horizon. The calculation of the QALY is straightforward when the values of health related quality of life (HRQoL) associated with health outcomes become available: the QALY is calculated as individual's length of life weighted by a valuation of their HRQoL over that period (Drummond et al., 2005). The valuation of the HRQoL consists of two parts: (1) the description of changes in HRQoL as a result of treatment, and (2) a valuation of that description of HRQoL. However, there are many health conditions treated by the NHS, and the description of quality of life might consider different aspects of individuals health. Thus, it is not possible to conduct a survey that values all possible changes in HRQoL for all health conditions. Therefore, changes in health/HRQoL need to be measured using a particular health instrument that reduces the number of health conditions evaluated so as to make the survey manageable, while being able to value all changes in health using a statistical model. In health economics, the change in HRQL is described using different multi-attribute health status classification systems such as European qality of life with five dimensions/attributes (EQ-5D), short form with three dimensions (SF-3D) (developed in ScHARR, Brazier et al. 2002) that can used for any illness, and other systems developed for specific conditions (e.g. AQL-5D for asthma).

Different classification systems produce different quality weights (utility scores), and hence results from different systems are not always comparable. Given that comparability is important for policy decision makers such as NICE, a single classification

system should be used for the measurement and valuation of HRQoL. Based on the comparative nature of NICE work and the need for consistency across appraisals, NICE prefers the EQ-5D measure from the EuroQol Group for preference measure. The EQ-5D consists of five dimensions/attributes of health: mobility, ability to self care, ability to perform usual activities, pain and discomfort, and anxiety and depression. Each attribute in turn consists of three levels ordered from less to most severity. Each health state is described in the form of a five-digit code using the three levels. For instance, the EQ-5D health state 11232 indicates no problems with mobility and self care, some limitation in the usual activities with extreme pain, and a moderate level of depression.

There has been some argument around the applicability of using a single preference measure for all interventions and patient groups, since some generic measures have been found to be sensitive or lacking in relevance to the conditions (Brazier and Tsuchiya, 2010). In the circumstance where EQ-5D is not an appropriate measure to describe the change in HRQoL, NICE requires empirical evidence to illustrate why it is not appropriate, and how the choice of other instrument would impact on the valuation of the QALYs.

Using these classification systems, health conditions can be then mapped to those health states defined by the underlying classification system, such as the EQ-5D, and hence be able to estimate the quality value of any conditions. The question now is how to elicit the quality values of these health descriptions (HRQoL), i.e. the 'Q' part of the QALY or also known as utility related to the HRQoL, in order to be used in computing the QALY values. These values are elicited directly from patients or general public using a choice based method. Two commonly used choice methods to value health outcomes (here, health states defined by the underlying classification system) are the time trade-off (TTO) and standard gamble (SG) methods. In the TTO technique, respondents are asked to choose between living for t years in their current health state (e.g. EQ-5D health state 11232) or living for x years in full health (EQ-5D state 11111), where $x < t$. The SG method captures the risk attitude, where

respondents have to make a choice between a certain outcome (e.g. EQ-5D health state 11232) and an uncertain outcome that has two possibilities; either return to full health with probability P or immediate death with probability $1-P$. Both evaluation techniques involve questions that might be complicated and contain biases as measures of preference (Brazier et al., 2007). Recently there has been increased interest in using an easier preference elicitation method to value health states within QALY framework: discrete choice experiment (DCE).

The methodology of deriving individuals' preferences/utilties using DCEs has been developed in market research since it was first introduced in the early 1970s (Louviere et al., 2000). The choice experiment involves asking consumers to choose the preferred product from a set of hypothetical products called a choice set. Each product is described by a combination of attribute levels which is called a profile. The technique is used in marketing to identify the importance of each characteristic of the product (attribute and its level) based on the specified preferences. This can tell economists how to improve the product based on consumer preferences, and hence maximise sales (Carson et al., 1994). In health economics, DCEs have been used recently to value health states utilities within the QALY framework. Patients are assumed to value different health states based on their attribute levels defined by the classification system under study. For instance, using the EQ-5D classification system a respondent might be asked to choose the preferred health state (profile) from the following choice set $\{11232, 12321\}$, where more preferred health states have higher health-related utility values.

Utilities are required for all health states defined by a classification system to be used within a health economic evaluation study. Nevertheless, it is not feasible to directly value all these health states using DCE, since a classification system might produce hundreds or thousands of health states which results in large choice experiments. For example, EQ-5D has 243 possible health states, and considering a DCE of pairwise comparisons, i.e. each choice set consists of two health states/profiles, this only would

require a valuation of 29,403 possible pairwise comparisons, which is infeasible. Because participants can only value a limited number of choice sets, rarely exceeding sixteen in health economic evaluation studies and more usually around eight (Ryan and Gerard, 2003). Therefore, a selection of those health states and choice sets should be evaluated. This collection of choice sets presented to individuals constitutes the design of the experiment. Accordingly, the choice of health states to be evaluated is essential if an efficient choice design is to be generated.

In most health economic evaluation studies, these health states are typically selected based on a simple design such as an orthogonal design (de Bekker-Grob et al., 2010), although generating choice experiments for health economic evaluation is a much more complex problem that requires more than just standard design methods. This is because, in addition to the basic choice of health states/profiles, consideration must be given to excluding unrealistic health states (e.g. health states with a combination of serious mobility health problem and no limitation in self care) as well as other design considerations for valuing health utilities within QALY framework.

1.4 Motivation for Better Choice Design

The main issue with choice studies conducted to value health state utilities is related to the design methods used to construct the DCEs. A recent overview of DCEs in health economics by de Bekker-Grob et al. (2010) shows that despite the advanced development of DCEs in different areas outside health economics, particularly in terms

of experimental designs and methods of analysis, the construction of the choice designs for valuing health state utilities is still rudimentary, such as using existing orthogonal rays design in SAS or SPSS programmes (e.g. the DCE study conducted by Brazier et al. (2009) to value asthma health states). The orthogonal designs, usually chosen for convenience, are based on linear principles, thereby ignoring the nonlinear nature of the choice model, and this might reduce the efficiency of the choice design. Therefore, we need to improve the methodology of constructing choice designs to produce more reliable assessments for health-related utility values.

Recently there has been a considerable development concerning the experimental design for choice studies, particularly within optimal design theory. There are algorithms for producing choice experiments that are optimal for different statistical measures (e.g. work by Kessels et al. 2006, 2008, 2011b). However, these developments have not received much attention from researchers in health economics. In addition, there is still further development required, for instance in terms of using more complex models that allow for taste variation in individual's preferences; and using prior assumptions about choice models' parameters, obtained from pilot studies, in the design development (Rose and Bliemer, 2008). This thesis employs the latest advanced work in design theory, while incorporating available information about both individuals' preferences and health state instruments, obtained from prior studies, in generating the choice design particularly for paired comparisons. We will use Bayesian optimal design theory, and aim to improve the statistical efficiency of the choice design and consequently the reliability of health state utility values.

Developing the survey design of health study is important to obtain precise estimates for the QALY values, and consequently improve the accuracy of the overall decision made using the views of the general public.

1.5 Aim of the Thesis

The use of DCEs for evaluation health states within the QALY framework seems promising. However, this technique still needs considerable development to produce reliable assessments for health state utility values. In this thesis, we consider how to improve the efficiency of choice designs for valuing health states. Our aims are the following.

1. Search the literature for the latest advanced work in discrete choice experiments, particularly within optimal design theory. The search will cover algorithms and software used to construct choice designs in different fields. We will investigate the ability of available software to construct a choice design for health evaluation studies.

2. Develop an efficient methodology for generating optimal or near optimal choice designs for health state utility determination within the QALY framework.

3. Provide different methods to assign an appropriate prior distribution for the choice model's parameters. Since constructing choice designs for non-linear models, here discrete choice models, depends on the unknown model's parameters where Bayesian approach will be used to overcome the dependency problem.

4. Investigate the effect of the choice of prior distribution on the design efficiency, and assess the robustness of the choice design to the choice od prior distribution.

1.6 Thesis Outline

The thesis consists of three parts. Part one provides a background knowledge of the main concepts of health economics and measuring health outcomes using QALYs, alongside a description of health state evaluation techniques, particularly using DCEs and their models (**Chapter 2**). This is followed by an overview of the current applications of DCEs in health economics, particularly those used for valuing health state utilities, and a literature review for the latest work in the optimal design theory for nonlinear models (specifically discrete choice models) in other economic areas (**Chapter 3**).

The second part of the thesis (**Chapters 4–6**) describes the application of a DCE conducted to value health states defined for asthma using the AQL-5D health instrument, and presents methods to produce and improve choice design for health evaluation studies.

- **Chapter 4** presents two real health economic studies conducted to value asthma health states utilities using the TTO and DCE techniques. It also provides an analysis for the choice data collected from those studies in both classical and Bayesian manner to obtain appropriate prior distributions for the construction of Bayesian choice designs.

- **Chapter 5** presents the software noted in the literature that generates choice designs for nonlinear models, and our attempt to construct choice design for the AQL-5D instrument using such software. In addition, it illustrates problems and difficulties in using this software to generate efficient choice design for health valuation purpose. And finally, it proposes our design algorithm to con-

struct choice design, particularly paired comparison design for the logit model, for valuing health states, taking into account health design requirements during the design phase. Thus, our design algorithm considers optimising the correct design criterion that takes into account the including of the death state in the choice model, i.e. the logit model, to anchor health state utility values within the QALY scale, as well as excluding unrealistic and dominant states from the optimal choice design.

- **Chapter 6** investigates the effect of using the prior information about the individuals' preferences on developing the efficiency of the design choices, and studies how a different choice of prior distribution could affect the choice of Bayesian optimal design.

The final part is a discussion of the results obtained, and presents a number of design limitations and recommendations for future work (**Chapter 7**).

Chapter 2

Valuing Health and Discrete Choice Experiments

2.1 Introduction

In health economic evaluation, it is essential to have a common health effect measure to be able to make a decisions about resources allocation across different health programmes or treatments which usually have different primary outcomes. Therefore, many health care organisations, such as NICE, recommend the use of cost-utility analysis (CUA). This type of health economic evaluation allows for comparisons across various health programmes by using a single health effect measure called the QALY that considers both the quality and quantity of life gained.

This chapter begins by defining the term QALY, and discusses how QALY are determined and used in CUA. This is followed by classifying the change in health-

related quality of life using different health descriptive systems in Section 2.3, and presenting methods to value these changes in health, particularly using discrete choice experiments (DCEs), discussed in Section 2.4. The final section of this chapter discusses modelling discrete choice data, and illustrates the derivation of a widely used discrete choice model, namely the multinomial logit model (MNL).

2.2 QALYs

In health economics, an extremely useful innovation has been developed to assess the benefits of different interventions in terms of health-related quality of life (HRQoL) and survival (in years), using a single measure called quality-adjusted life years (QALYs). Weinstein et al. (1996) define the QALY as a measure of health outcomes which assigns to each period of time a weight, ranging from 0 to 1, corresponding to the health-related quality of life during that period, where a weight of 1 corresponds to optimal health and a weight of 0 corresponds to health state judged to be equivalent to death.

The number of QALYs associated with a health outcome (e.g. health state) is expressed as time of life spent in a specific health state weighted by a valuation of that state. If an individual is expected to live Y years in less than full health, then the number of QALYs experienced is equivalent to living X years in perfect health where $X < Y$ (Brazier et al., 2007). Thus, a year of perfect health is equivalent to one QALY. This could be divided between several individuals or years. For instance, one QALY is equivalent to four people experiencing one year in a health state valued at 0.25, or one person living for two years worth 0.5 QALY weight.

The weights, also known as utilities, should be based on a preference measure such

that the more desirable health state receives more weight, and hence will be preferred in the analysis. They also must be measured on an interval scaled relative to perfect health and death. These two points are required since they both occur in any health programme evaluated using QALY analysis, and weights will be required for them. In the literature the most convenient scale for the utility scores is the 0–1 scale, where zero represents death and one reflects perfect health (Drummond et al., 2005). However, it is still possible to assign negative values for states that are worse than death, and values more than one for states better than perfect health if they exist.

The QALY weights are used in the CUA to determine the most cost effective treatment for the NHS budget. This can be illustrated by considering the following example.

2.2.1 Illustrative Example

In this section we illustrate the way of computing QALYs gained from different treatments and their uses in CUA to determine the most cost effective treatment.

Suppose there are two treatments to be considered by the NHS to recover from back pain. Treatment 1 is a new drug that costs £10,000 per patient, and is expected to extend patient life for 4.25 years (4 years 3 months) with a quality of life less than perfect worth 0.6. Treatment 2 is the standard care that costs £3,000 per patient, and receiving such treatment is assumed to generate 4 additional years in a health state valued at 0.55.

Suppose that the NHS has to prioritise funding for one treatment, so that a choice has to be made between the treatments. In CUA, such a decision is made based on comparing cost effectiveness of both treatments using the ICER, to decide whether the extra cost of the new drug is worth the small changes in the quality of life. The ICER can be calculated as defined in equation (1.1.1),

$$\text{ICER} = \frac{\Delta C}{\Delta E} < \kappa,$$

where ΔC and ΔE are the mean differential costs and effects (population mean number of QALYs gained), respectively. Treatment is said to be cost effective if ICER is less than κ, which is a predefined value of willingness to pay per QALY gained.

Now, to compute the cost effectiveness of the provided treatments, we need to compare the additional cost of the new drug, $£7,000$, to its extra effect compared to the standard treatment (QALYs gained). Comparing the new drug with the standard treatment in terms of QALYs gained indicates that the new drug leads to an additional 0.35 QALY – that is, $\text{QALY}_G = \text{QALY}_{T1} - \text{QALY}_{T2} = (4.25 \times 0.6) - (4 \times 0.55) = 2.55 - 2.2\text{QALYs} = 0.35$ QALY. Thus, the ICER $= \frac{£7000}{0.35} = £20,000$ per QALY, so the new drug would cost an additional $£20,000$ per additional case successfully treated. Using a threshold value of $£30,000$, for example, for κ indicates that the new drug is cost effective and provides good value for NHS budget, although its changes to the quality of life compared with the standard care is small.

The determination of the quality weight or the 'Q' part of the QALY needs a measure to describe the effect of disease or its treatment in the HRQoL, as explained in Section 2.3, and a technique to value these description of health consequences, which will be discussed in Section 2.4.

2.3 Classification Systems

The quality of life experienced during a specific time covers a whole range of different aspects of individual's health, including physical state, mental capacity and social activities, and hence not only the absence of disease. This multidimensional definition of health or HRQoL and the variety of illnesses treated by the NHS make it impossible to

assign a QALY value to each health condition directly. Therefore, a health instrument is needed to define and value a finite set of health states while being able to estimate the QALY values for all possible health conditions. The health instrument describes the impact of diseases or their treatment on HRQoL in terms of the most important features related to a health condition (e.g. symptoms, general well-being). These features are called attributes. For each attribute various levels of severity are defined, for example, no problem, little problem, and extreme problems. A set of attributes and levels constitute what is called the multi-attribute health status classification system, and a combination of the attribute levels defines a health state.

Those health classification systems enable health economists to assign QALY values for all possible health conditions. This is possible through obtaining QALY values for a subset of health states defined by a classification system, then estimating the QALYs for all health states defined by the system using a statistical model, as described in Section 2.3.3. QALY values are then obtained for any real world state of health by mapping this state on to the classification system.

Health economists discriminate between two types of health instruments based on their contents: generic classification systems general to any health condition, and condition specific classification systems more specific for disease symptoms. A description and examples for each type are provided in Sections 2.3.1 and 2.3.2, respectively.

2.3.1 Generic Classification System

The generic classification systems, used for any type of population, cover general characteristics of health such as mobility, pain, activity limitation and depression. They do not cover small and important aspects of specific conditions or diseases as this may make them inappropriate for all conditions. In health economics, there are many generic health instruments, such as the EQ-5D, SF-6D and health utility index (HUI) In this section, we review the widely used EQ-5D instrument.

- **EQ-5D**

  The European Quality of Life (EUuroQoL) group was initially developed by EuroQol (1990) with six attributes, and then revised to include five attributes: mobility, self care, usual activity, pain/discomfort, and aniety/depression (Brooks, 1996). Each attribute has three levels of severity ordered from no problem (level 1) to major problem (level 3). This produces $3^5 = 243$ possible health states which is raised to 245 states when dead and unconscious are added for scaling purpose. The perfect health state is defined as the combination of the best level of each attribute. Each health state is described as a five-digit code using the three levels. For instance, health state 11222 indicates no problem with mobility and self care, some limitation with usual activities, and moderate pain and level of depression.

NICE recommends measuring HRQoL, and hence QALYs, using the EQ-5D to ensure that all patients in different conditions are being assessed against the same health features. This allows for the comparison of health interventions with different primary outcomes (NICE 2008). NICE argues that the generality of this measure makes it applicable for all interventions and groups of patients. However, this instrument has been shown to perform poorly with some conditions, such as in visual impairment in macular degeneration (Espallargues et al., 2005), hearing loss (Barton et al., 2004), and leg ulcers (Walters et al., 1999), as its attributes focus on general rather than specific aspects of health. An alternative is to use a more specific descriptive system that captures the impact on HRQoL of patients with specific diseases. In this case, NICE requires the provision of evidence to explain why the EQ-5D health status system is not appropriate for a specific class of patients, and clarification of the use of the new instrument.

2.3.2 Condition Specific Classification System

Condition specific health state instruments are used with specific populations who have a particular condition or disease. The health state utilities obtained using condition specific preference-based measures (CSPMs) might produce more relevant economic evaluations than generic preference-based measures (GPMs) (Yang et al., 2010), as they are more relevant to changes in HRQoL associated with a specific illness. An example of such an instrument is the Asthma Quality of Life classification with five dimensions/attributes (AQL-5D), shown in Table 2.1. This health instrument will be used for constructing and producing the analysis of the choice designs throughout this thesis.

- **AQL-5D**

  The Asthma quality of life (AQL-5D) is a specific descriptive system derived from the condition specific instrument AQLQ. It is designed to describe HRQoL in adult patients with asthma. Initially, the AQLQ consists of 32 items covering four dimensions of health: asthma symptoms (12 items), activity (11 items), emotional function (5 items), and environmental stimuli (4 items). Each item has seven levels, ranging from no problems to extreme problems. This numerous number of health dimensions and levels produces millions of possible health states, where each health state involves a considerable amount of information to be evaluated by respondents. This complicates the valuation process, as respondents have difficulty in valuing health states with more than 9 attributes (Brazier et al., 2012).

  To simplify the valuation of HRQoL described by the AQLQ instrument, Yang et al. (2007) developed an approach to produce a health state classification system from the large AQLQ instrument with five attributes only: concern about asthma, shortness of breath, effect of weather and pollution, the impact of asthma on sleep and general activities. Each attribute has five levels of severity as shown

in Table 2.1, where level 0 is used to indicate no problem and level 4 for extreme problems, defining 3,125 possible health states. For example, asthma health state 11244 indicates that a patient feels concern about having asthma and short of breath a little of the time, his/her health is affected by the weather and pollution sometimes, and he/she cannot have a good night's sleep, and has extreme limitations in all activities.

Table 2.1: Asthma Quality of Life Classification System (AQL-5D)

Attributes	***Attribute Levels***
Concern	Feel concerned about having asthma none of the time. Feel concerned about having asthma a little or hardly any of the time. Feel concerned about having asthma some of the time. Feel concerned about having asthma most of the time. Feel concern about having asthma all of the time.
Short of Breath	Feel short of breath as a result of asthma none of the time. Feel short of breath as a result of asthma a little or hardly any of the time. Feel short of breath as a result of asthma some of the time. Feel short of breath as a result of asthma most of the time. Feel short of breath as a result of asthma all of the time.
Weather and Pollution	Experience asthma symptoms as a result of air pollution none of the time. Experience asthma symptoms as a result of air pollution a little or hardly of the time. Experience asthma symptoms as a result of air pollution some of the time. Experience asthma symptoms as a result of air pollution most of the time. Experience asthma symptoms as a result of air pollution all of the time.
Sleep	Asthma interferes with getting a good night's sleep none of the time. Asthma interferes with getting a good night's sleep a little or hardly any of the time. Asthma interferes with getting a good night's sleep some of the time. Asthma interferes with getting a good night's sleep most of the time. Asthma interferes with getting a good night's sleep all of the time.
Activities	Overall, not limited with all the activities done. Overall, a little limitation with all the activities done. Overall, moderate or some limitation with all the activities done. Overall, extremely or very limited with all the activities done. Overall, totally limited with all the activities done.

These classification systems produce hundreds or thousands of health states, and this number increases as the number of attributes and levels increases. The main questions now are how to evaluate and estimate the utility value for all health states defined by a classification system, and which of these health states should be presented to respondents?

2.3.3 Modelling Health State Classification System Valuation

Eliciting the utility values, directly, for all health states defined by a classification system is not practical, since there are too many health states produced by a classification system (e.g., AQL-5D produces 3,125 health states). The solution to this difficulty is to value a selection of health states using one of the preference measures illustrated in Section 2.4 (e.g., time trade-off or standard gamble methods), then estimate a model for predicting the utility values for all health states defined by that classification system.

A range of models have been developed to fit and analyse the elicited preference data, and estimate health state valuations. Here, we consider the fundamental statistical model used to estimate health state values. The basic model defines health state utility as a function of that state, that is the attributes and attribute levels of the classification system. This is typically expressed as in McFadden (1974) by equation (2.3.1), where the latent utility of individual i valuing health state $\boldsymbol{x}_{ij}$ is decomposed into two parts: a systematic component, $g(\boldsymbol{x}_{ij})$, defined as the population mean utility, which is a function of the attributes that make up the states; and a random component, ϵ_{ij}, that represents the variation around the population mean utility.

$$U_{ij} = g(\boldsymbol{x}_{ij}) + \epsilon_{ij}. \tag{2.3.1}$$

In this thesis, the population mean utility, $g(\boldsymbol{x}_{ij})$, is defined as a linear additive model of the attribute levels:

$$g(\boldsymbol{x}_{ij}) = 1 - \boldsymbol{\beta}\boldsymbol{x}_{ij}^T, \tag{2.3.2}$$

where $\boldsymbol{\beta}\boldsymbol{x}_{ij}^T$ represents the utility loss from perfect health, which is here mapped as the best health state defined by the classification system, to health state $\boldsymbol{x}_{ij}$. We write the

health state $\boldsymbol{x}_{ij}$ as a vector of dummy variables with elements defined as

$$x_{\lambda\delta} = \begin{cases} 1 & \text{if attribute } \delta \text{ of health state } \underline{x}_{ij} \text{ is at level } \lambda \text{ or higher,} \\ 0 & \text{otherwise.} \end{cases}$$

For example, in terms of the AQL-5D classification system, where each health state is defined by 5 attributes each with 5 levels of severity, health state $\boldsymbol{x}_{ij}$ would be a vector of 20 dummy variables; $\boldsymbol{r}_{ij} = (x_{11}, x_{21}, \ldots, x_{41}, \ldots, x_{15}, \ldots, x_{45})$. For instance, health state 13402 would result in $\boldsymbol{r}_{ij} = (1,0,0,0,1,1,1,0,1,1,1,1,0,0,0,0,1,1,0,0)$.

Health state valuations must follow the convention that utilities are defined relative to the utilities of perfect health and death, where perfect health has a utility of 1 and death has a utility of 0 as discussed in Section 2.2; therefore, a dummy variable corresponding to death, x_d, is included in the representation of health state utility, where

$$x_d = \begin{cases} 1 & \text{if health state } \boldsymbol{x}_{ij} \text{ is the death state,} \\ 0 & \text{otherwise.} \end{cases}$$

Thus, health state 13402 would be represented by a vector of 21 dummy variables with $x_d = 0$, and the death state is represented by a vector where the first 20 elements are 0, and the last element is 1.

The first 20 elements of the corresponding vector to the unknown parameters of the utility model, $\boldsymbol{\beta} = (\beta_{11}, \ldots, \beta_{41}, \ldots, \beta_{15}, \ldots, \beta_{45}, \beta_d)$, represent the decrease in utility associated with moving one level on one attribute, and β_d represents the decrease in utility associated with moving from perfect health to immediate death. Thus, the mean utility value of any health state is computed as 1 minus the sum of the coefficients corresponding to the attribute levels defining the health state. However, based on this definition, the utility value of death is not anchored at zero, so the estimated utility values are not anchored on the 0-1 scale required for calculating the QALYs. Following the rescaling method in McCabe et al. (2006) of dividing all parameters of the proposed health state utility model by the death coefficient, that is $\beta^r_{\lambda\delta} = \beta_{\lambda\delta}/\beta_d$, ensures that

death has zero utility.

According to this definition of the population mean utility, $g(\boldsymbol{x}_{ij})$, perfect health would have utility of 1, since the dummy variables in $\boldsymbol{x}_{ij}$ are all zero. For the death state, $\boldsymbol{\beta}\boldsymbol{x}_{ij}^T = 1$ and hence the mean utility value is defined to be 0. This follows since for death state the dummy variables in $\boldsymbol{x}_{ij}$ are all zero except the death variable, and the parameter of death $\beta_d = 1$. Using this statistical model, utility values can be estimated for any state defined by the AQL-5D classification system within the required scaled of the QALYs, while retaining the possibility of having health states worse than death (i.e. $U_{ij} < 0$).

2.4 Measuring Preference

Having introduced different classification systems to describe health conditions, and modelled their health state valuations; it is essential to present how these values or health state utilities are elicited for a subset of health states. In this section, we review various valuation techniques used in health economics for measuring health state utility that reveal individuals' preferences for particular health states; time trade-off and standard gamble methods are discussed in Section 2.4.1, and some alternatives in Section 2.4.3.

2.4.1 Direct Valuation Techniques

Direct valuation techniques, also know as cardinal techniques, are used to measure individuals' preference of health state for particular health conditions. In such tech-

niques, a health state is assigned a weight/value reflecting the strength of preference of this health state relative to the worst and the best defined health states, often death and full health, respectively. Thus, cardinal values are anchored directly on the 0–1 scale. In this section we review the most widely used techniques; time trade-off (TTO), standard gamble (SG), and visual analogue scale (VAS).

2.4.1.1 Time Trade-Off

The time trade-off (TTO) technique was developed by Torrance et al. (1972). In this method, respondents are asked to trade off between the improvement in their quality of life and the number of life years they are willing to sacrifice in order to avoid a certain poorer health state (Brazier et al., 2007). For health states considered better than death, the respondent is asked to choose between two certain options:

1. Living for t years in health state i, worse than perfect health, followed by death.
2. Perfect health for time $x < t$ years followed by death.

The time x is varied until the respondent is indifferent between the two choices, for which the utility of health state i is given by $\frac{x}{t}$.

If health state i is considered worse than death, the respondent will be given two alternatives:

1. Health state i for time $t - x$, where $x < t$, followed by perfect health.
2. Immediate death.

Again time x is varied until the respondent is indifferent between the two alternatives. The utility for health state i is then defined as $\frac{-x}{t-x}$.

- **Example:**

  Consider the TTO exercise in Yang et al. (2007) that interviews 300 individuals

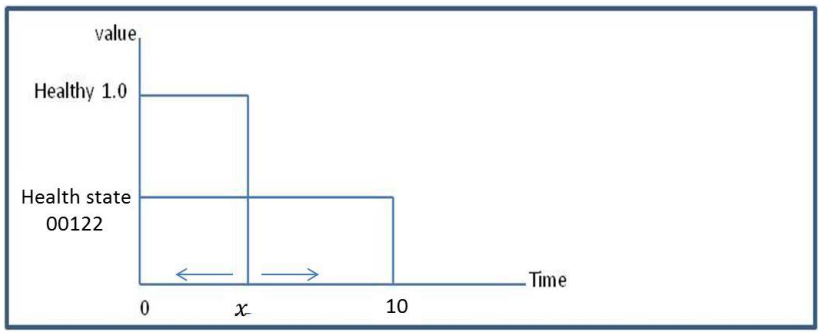

Figure 2.1: Time-trade off (TTO) technique for asthma health state preferred to death, redrawn from Brazier et al. (2007)

to elicit utility values for 98 asthma health states. Considering asthma health state 00122, an individual is asked to imagine being in AQL-5D state 00122 for a remaining life expectancy of 10 years; and then he/she is provided with two options: either living for 10 years in state 00122 or living shorter time, $x < 10$, in perfect health (the best health state defined by the AQL-5D, 00000) followed by death, as shown in Figure 2.1.

The individual is then asked to consider a number of shorter periods in perfect health that makes him/her indifferent between choices. Now as the provided health state is pretty good, the individual may not be willing to trade much time. The study shows that the health state utility value of an individual is 0.829 QALYs – that is, the individual is unable to choose between health state 00122 and being healthy when $x = 8$. The task is then repeated for several individuals for all the selected health states. The elicited TTO values for each health state can be used to estimate utility model parameters using individual level models, which considers variation between respondents (e.g. a random effect model), or aggregate level models that are estimated based on the mean TTO values of each health state.

2.4.1.2 Standard Gamble

The standard gamble (SG) is a classical method to measure utility. In the SG method, respondents are faced by two choices: a certain health state and a risky option. The method varies, depending on whether the health state is preferred to death or is considered worse than death. For a health state considered better than death the respondent is offered two alternatives:

1. Treatment with uncertain outcomes: either return to perfect health and live for an additional t years then death, with probability P; or immediate death with probability $1 - P$.
2. Health state i with certainty for t years followed by death.

The probability P is varied until the respondent is indifferent between the two alternatives. The utility value of health state i for t years is equal to the given probability P for the better outcomes in the risky option.

For a health state considered worse than death, the participant is shown two alternatives:

1. Treatment with uncertain outcomes of perfect health for t years with probability P, and health state i with probability $1 - P$ of living for t years, again both followed by death.
2. Immediate death.

The probability P is varied until the respondent is indifferent between the two alternatives. In this case, the utility value of health state i is given as the negative value of the probabilities ratio, $\frac{-P}{1-P}$.

- **Example:**

  Consider the example of being in asthma health state 00122 for 10 years followed

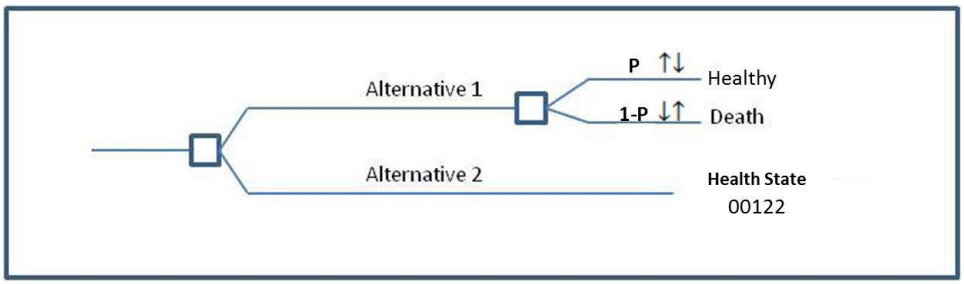

Figure 2.2: Standard gamble technique for asthma health state preferred to death, redrawn from Brazier et al. (2007)

by death, where here time is consider to be fixed). The SG technique asks a different type of question. In this method, respondents would be provided with two alternatives; living for 10 years in asthma state 00122, or taking a risky option with two uncertain outcomes: either living an additional 10 years in perfect health (the best health state defined by the AQL-5D, 00000) or immediate death, with probabilities P and $1 - P$ respectively, as illustrated in Figure 2.2.

Individuals are then asked to provide the probability value that makes them indifferent between the two alternatives. A probability wheel is typically used in the SG task to help individuals to elicit the probability value. The wheel is divided into two parts with a different colour representing each outcome of the risky option: death and full health. Equal parts would show a 50/50 chance of receiving either outcome of full health or death. Given this probability, the asthma state may be more attractive for respondents than the risky option. Thus, the interviewer would keep increasing the probability of perfect health in the probability wheel until respondents are indifferent between the two alternatives. Now, suppose an individual chose 0.8 and 0.2 as the probabilities for perfect health and death, respectively, that make both choices equally attractive. The choice task will be done, and the utility value of state 00122 is equal to 0.80.

2.4.1.3 Visual Analogue Scale

In the visual analogue scale (VAS) technique, which is also called a rating scale, respondents are asked to value health states on a scale that is usually arranged from 0 to 100. An individual is first asked to rank a given set of health states from most preferred (represents one end of the scale) to least preferred (represents the other end of the scale). The health states are then allocated relative to each other on a scale such that spaces between health states correspond to the difference in preference as perceived by the subject. Therefore, health states with similar preference should be placed close to each other, whereas health states that are very different in desirability would be placed far apart (Drummond et al., 2005, pp.147–149).

To measure preference for health states using the rating scale method, individuals must consider every given health state as a permanent state, and they all last for the same time and are followed by death. If the individual chooses perfect health and death as the most and the least preferred states, respectively, the utility value for health state i would be placed between these two states, represented by the value associated with its placement x (Brazier et al., 2007). In the case where death is not allocated as the worse health state, the utility value of health state i is defined as $\frac{x-d}{1-d}$, where x and d are the values corresponding to the placement of health state i and death, respectively.

In practice, the VAS method is often used as a warm-up exercise before performing a valuation technique such as SG and TTO methods. This allows respondents to become more familiar with comparing health states, and hence value health more accurately (Brazier et al., 1999). Further descriptions of these techniques can be found in Brazier et al. (2007) and Drummond et al. (2005, pp.147–153).

2.4.2 Direct Valuation Method Issues

In this section, we discuss the empirical issues of the cardinal methods, particularly the TTO and SG methods. In general, different valuation techniques may produce different utility values for the same health state, and health economists mainly advocate the use of choice-based methods such as TTO and SG over the rating methods, for example, the VAS method Tolley (2009). However, both the TTO and SG techniques have been criticized by many researchers for the difficulty some respondent groups have in understanding them (e.g., Brazier et al., 2007 and Flynn, 2009), as cognitive ability varies across individuals and might be limited in particular groups of the population. The SG method, for example, requires some understanding of the probability concept in order to perform the task, and this might be limited in an uneducated population. Thus, an interview is usually required to administrate such a task, and that might be time consuming and expensive.

In addition, the utility value estimates using these methods are affected by other non-health factors such as time preference (longevity) in the TTO method and risk attitude in SG methods (Brazier et al., 2007). The different effects of these factors on the individuals' preferences produced by SG and TTO techniques is explained by Bleichrodt (2002). He identifies four possible sources of biases in these methods which could move the utility values upward or downward depending on the technique used to measure the preferences: utility curvature, probability weighting, loss aversion and scale compatibility. For example, utility curvature does not lead to bias in SG utility values, since the utility is not restricted on duration of health states, whereas the TTO utility function is assumed to be linear in duration. Thus, it is expected that most respondents would show positive time preference that leads to upward bias in TTO values. The opposite is observed for the probability weight, as it affects the SG utility values only.

Additionally, both techniques are limited in their ability for valuing health benefits

beyond direct health outcomes. Thus, in general, for studies that consider valuing indirect health care benefits such as non-health outcomes (e.g, provision of information, and reassurance) and process attributes (e.g., treatment location, and route of drug administration) the use of a cardinal method is not appropriate.

The limitations of these measurement methods in their cover of non-health outcomes and their complexity has lead to an increase interest in seeking alternative measurement techniques. Recently there has been increased interest in improving the choice technique so that it becomes simpler and easier for respondents to use. Some alternative techniques are illustrated in the following section, with an emphasis on the discrete choice experiment method.

2.4.3 Indirect Valuation Techniques

Indirect valuation methods are an alternative technique to elicit preference for health outcomes of particular interventions. They are also know as ordinal methods. In such techniques, health outcomes typically need to be rank ordered to allow the selection of the most preferred option. There are several techniques for eliciting ordinal values for health states. We describe the ranking and discrete choice experiments (DCEs) methods in Sections 2.4.3.1 and 2.4.3.2, respectively.

2.4.3.1 Ranking

In a ranking task respondents are asked to order a set of health states from the best to the worst state (Brazier et al., 2007). The ranking method is typically used as a warm-up task rather than as the main method for deriving health state utility. It has been used in many health evaluation studies, such as Brazier et al. (2002) and Dolan et al. (1996a).

2.4.3.2 Discrete Choice Experiments

A discrete choice experiment (DCE), usually consists of several choice tasks. In each task, respondents are faced by two or more hypothetical health states called profiles or alternatives and asked to choose the most preferred state. Choices are then modelled to make inferences about the underlying utility function, which we discuss in Section 2.5.

- **Example**

  To illustrate the DCE technique consider the following paired comparison of asthma health states defined by the AQL-5D classification system from a study by Brazier and Tsuchiya (2006) to estimate the health state utility values for QALYs computation.

Health State A (12101)	Health State B (43220)
Feel concern about having asthma *a little of the time.*	Feel concern about having asthma *all of the time.*
Feel short of breath as a result of asthma *some of the time.*	Feel short of breath as a result of asthma *most of the time.*
Experience asthma symptoms as a result of air pollution *a little of the time.*	Experience asthma symptoms as a result of air pollution *some of the time.*
Not having a good night sleep as a result of asthma *none of the time.*	Not having a good night sleep as a result of asthma *some of the time*
Have *a little* limitation with all activities done.	Have *no* limitation with all the activities done.

Which health state would you prefer? A or B

In the choice task, respondents would be shown either a representation or a full description of health states, and then asked to select their most preferred health state. It is assumed that respondents will consider all the information provided to them and make their choices based on the maximum utility; the alternative with the highest utility will be chosen. For the provided pair-wise comparison, the study shows that health state A is more preferred to health state B with observed choice probability of 0.846 (33 out of 39 respondents preferred health

state A). The observed choices from the multi choice tasks in the DCE enable the researchers to model the probability of an alternative being selected as a function of the attributes and the socio-economic characteristics of the respondents, and then make inference about the underlying utilities.

Ranking a set of states could be regarded as a series of discrete choice tasks. For example, for ranking three health states A, B and C, a subject might treat it as a sequence of discrete choice tasks (Brazier et al., 2007). Thus, it might be considered as either a paired comparisons, A over B, and B over C, or as choices within subsets A from the set $\{A, B, C\}$ and B from $\{B, C\}$. This assumes "independence from irrelevant alternatives (IIA)" – that is, ordering any two health states is independent from the other alternatives available. This is also called Luce's choice axiom (Luce 1995).

Discrete choice experiments have been used widely in different areas since they were introduced in marketing in the early 1970s and have started to receive more attention from both academic and industrial fields. They have been used in different economic fields to explore consumers' preferences for attributes of goods and services, and to model consumers' decisions to predict future market demand (Carson et al., 1994). The technique was introduced to health economics in Propper (Propper, 1990) to estimate the cost that patients are willing to pay to reduce the waiting times spent in NHS for non-urgent care. Since then interest has been increased in using DCEs in health evaluation studies to value health outcomes and health care benefits beyond direct health outcomes, as health benefits might extend to include non-health outcomes, such as provision of information, reassurance, autonomy and dignity in the provision of care, and process attributes, such as treatment location, route of drug administration and patient experienced burden of testing (Ryan, 1999). More recently, there has been increased interest in using such techniques for valuing health outcomes (health states) within the QALY framework. In the following section, we briefly discuss the advantages and disadvantages of using such techniques to value health outcomes to

produce cardinal values or utility value within the QALY framework.

2.4.4 Advantages and Disadvantages of DCE Techniques

The main advantage of ordinal methods is that they are relatively easy to comprehend and administrate. Thus, they might be more appropriate in most applications rather than cardinal methods such as TTO and SG techniques, particularly in situations and populations where educational ability and numeracy are limited. Measurement errors are reduced and this increases the reliability of health state valuations.

Additionally, elicited preferences from ordinal techniques are not contaminated by other factors such as risk (as in the SG method) or time preference (as in the TTO method). This is because the choice task in the ordinal techniques is designed such that respondents are forced to make choices over health states by trading their attributes and attribute levels without the need for external factors. So preferences or utility values can be interpreted as pure valuation of health states.

The main issue with health state utility values inferred using discrete choice data is that they are not directly anchored on the death and perfect health scale required for the QALY calculation (Flynn, 2010). Therefore, they cannot be used immediately for estimating the QALYs, or consequently in CEA or CUA that use cost per QALYs analysis. In the health economic literature, there have been several suggestion to overcome this problem.

1. Re-scale the coefficients on a latent utility scale using the TTO value of worst health state defined by the classification system (Ratcliffe et al., 2009). The DCE value for the worst state is anchored at the elicited TTO value of the worst state. Anchoring DCE values using this method has been criticized by many researchers (e.g., Bansback et al., 2010 and Rowen et al., 2011), as this method depends on an external cardinal measure, TTO technique, and this contradicts with the motivation of using DCE as an alternative for the conventional methods.

2. Include survival as a separate attribute in the DCE task. The resulting DCE would resemble the TTO exercise, as respondents choose between health states based on their description and the length of life provided for each health state. However, this might increase the complexity of the choice task for respondents. Also, this anchoring method complicates the design and modelling problems, as discussed by Bansback et al. (2010), since survival has a multiplicative (not additive) relationship to health-related quality of life in the QALY model.

3. A more recent method is to include death state comparisons in the choice design; this allows the estimation of the death coefficient and the scaling of the parameters of the latent utility model relative to the death parameter, as discussed in Section 2.3.3. This method has been investigated by Brazier et al. (2009). The study concludes that problems may only arise when many respondents do not regard any state defined by the classification system as worse than being dead, and so effectively not be willing to trade.

4. There are two new methods proposed recently in Rowen et al. (2011) that show improvement in anchoring DCE values for the AQL-5D states onto the 0–1 QALY scale, and predicting the mean TTO value of AQL-5D health states usign the DCE data. These are mapping and hybrid models methods. The first method is based on mapping the DCE values on a latent scale onto the TTO values using the simple mapping function

$$TTO_j = f(DCE_j) + \epsilon_j,$$

where TTO_j and DCE_j represent the mean TTO value and the latent utility value of state $\boldsymbol{x}_j$ modelled using DCE data, respectively, and ϵ_j is the error term. The second method anchors the DCE latent utility values by estimating the utility model coefficients by analysing both DCE and TTO data at individual level using a hybrid model. However, both anchoring methods require the TTO values for a subset of health states defined by the underlying classification system,

which again conflicts with establishing the DCE as an alternative for the cardinal methods.

DCEs seem to be a promising alternative to the cardinal techniques, which would reduce survey administration times and effort for collecting data. However, more work is still required to improve the data collected from this technique. In particular, further research is needed in terms of using more sophisticated design approaches to construct the DCEs (e.g. selecting health states presented to respondents and grouping them into choice tasks) while taking into account anchoring health state utility values and other design consideration for valuing health that will be discussed in Chapter 3. This is the key objective of our project.

An issue with health state valuations generated with these techniques is that different methods may produce different values, and not all generated values represent utilities in the formal sense described in O'Hagan et al. (2004, pp.41–46). From the preference methods reviewed above, only the SG method produces health state values that are considered as utilities (Drummond et al., 2005, pp.143–147). This method captures the risk attitude that is related to preference measurement and utility theory. Methods such as TTO generate a health state value that is not necessarily a formal utility. Ordinal techniques such as DCEs are assumed to produced utilities, since respondents are assumed to make their choice based on comparing different health states in their head, considering the level of severity provided for each health state; hence the most preferred health state is assumed to maximise their utility. However, as this thesis mainly considers the design phase of a choice study, the term 'utility' will be used to represent the value of a health state regardless the method used to generate it.

2.5 Discrete Choice Models

As we mentioned earlier in Section 2.3.3, the mean health state utility can be estimated for all health conditions through eliciting the utility values for a subset of health states defined by a classification system, and then estimating a model for predicting the utility value as discussed. However, utility values cannot be elicited directly from DCE techniques, as they provide choice frequencies of one chosen state over another, probabilities, rather than measuring the actual individual's preference of a specific health state (health state utility). Therefore, a method is needed to model the observed DCE data such that health researchers are able to make inferences about the parameters of the underlying utility model, and hence computing the utility and the QALY values for all health conditions.

In this section, we explain how to model DCE data related to the attribute levels that make up the states (mean utility) using random utility theory (RUT), and give an example of a widely used choice model, which is also the main model used throughout this thesis, namely the multinomial logit (MNL) model. The section begins with reviewing the basic concept of the random utility approaches, and then presents the identifiability issues associated with the discrete choice models (DCMs) in Section 2.5.2. Modelling health state valuations using the MNL model, and the derivation of the likelihood function are described in Section 2.5.3.

2.5.1 Modelling Discrete Choice Data

As we mentioned earlier, it is necessary to elicit health state utility for a subset of health states in order to estimate the preferences/weights associated with the attribute levels

defined the states, and consequently predict health state utility for all possible health states of particular conditions. However, utility cannot be elicited directly from DCE data, and therefore a method is needed to infer the parameters of the utility model from the discrete choice data.

The DCE data are modelled using different DCMs, such as conditional logit and probit models. The choice models are based on utility maximisation behaviour of respondents assumed by the RUT proposed by Thurstone (1927) in psychology and introduced to economics in Marschak (1960) and then developed further by McFadden (1974). The choice models basically relate the observed choices to the utility function defining the states, without reference to exactly how choice is made (Train, 2003, pp. 14-18).

In a choice experiment respondents are assumed to compare health states based on their attributes and attribute levels, where the most preferred state is assumed to have the highest utility. And the idea behind RUT theory in economic is that respondents obtain utility for each alternative in their head, and choose alternative that maximises their utility. This utility cannot be observed directly by researchers. The researchers actually observe the choices made based on the attributes and levels of the states beside other attributes of respondents (e.g., individual's income and socio-economic factors). Therefore, the researchers can specify a function to relate an individual's utility of any health state to these attributes, and then use DCM to infer the parameter in the utility function using the choice data. And since this thesis is considering the construction of efficient choice design to estimate health state utilities within the required QALY (0–1 scale), the utility function can be defined as in equation (2.3.1):

$$U_{ij} = g(\boldsymbol{x}_{ij}) + \epsilon_{ij},$$

where $g(\boldsymbol{x}_{ij})$ is the population mean utility defined by $g(\boldsymbol{x}_{ij}) = 1 - \boldsymbol{\beta}\boldsymbol{x}_{ij}^T$ as proposed in McCabe et al. (2006) to analyse the ranking data and discussed in Section 2.3.3, and ϵ_{ij} represents the variation around the population mean utility.

Under RUT, individuals are assumed to be utility maximisers, that is individual i will choose health state $\boldsymbol{x}_{ij}$ out of J alternatives offered in a choice set s, $C_s = \{\boldsymbol{x}_{i1s}, \ldots, \boldsymbol{x}_{iJs}\}$, if and only if $U_{ijs} > U_{its}$ for all $t \neq j$. Thus the probability of state $\boldsymbol{x}_{ijs}$ being chosen by individual i in choice set s is

$$\begin{aligned} P_{ijs} &= P\left[U_{ijs} > U_{its} \quad \forall t \neq j\right], \\ &= P\left[g(\boldsymbol{x}_{ijs}) + \epsilon_{ijs} > g(\boldsymbol{x}_{its}) + \epsilon_{its} \quad \forall t \neq j\right]. \end{aligned} \tag{2.5.1}$$

If the value of $g(\boldsymbol{x}_{ijs})$ is given for all j, then the choice probability p_{ijs} would depend on the distribution of the error terms $\epsilon_{i1s}, \ldots, \epsilon_{iJs}$. Different discrete choice models are obtained from different specifications of the joint density of the random errors.

The choice of random error distribution is considered as a starting point for selecting DCM models. The random errors could be assumed uncorrelated and identically type 1 extreme value distributed (i.e. having the same variance), which results in what is called the multinomial logit (MNL) model. This model is widely used in choice studies, though it has some restrictive assumptions: (i) independent and identically distributed (i.i.d.) errors, (ii) independence of irrelevant alternatives (IIA), which states that choosing between any two options is independent from other states provided in a choice set, (iii) homogenous preferences across individuals (i.e. preferences/weights of the utility model parameters is fixed over individuals). However, the popularity of this model is due to the fact that those assumptions allow a closed form for the choice probability, as will be shown in Section 2.5.3, which simplifies the calculation of the likelihood function as well as the constructing of the choice design. Nevertheless, they might be very restrictive in describing human choice behaviour. Therefore, other models have been developed from relaxing some of those assumptions and allowing for correlation between alternatives (relaxing the IIA assumption) as in the generalized extreme value (GEV) model, and individual's heterogeneity (heterogenous preferences/random coefficients across individuals) in the mixed nested logit (MNL) model (de Bekker-Grob et al., 2010).

The errors could also be assumed normally distributed with correlation or i.i.d normally distributed, which results in what is called a probit model. The probit model can also allow for individual's heterogeneity (random taste variation); this assumes that the utility model coefficients are normally distributed. The flexibility of the error correlation is the main advantage of this model, as it relaxes the first two assumptions of the logit model. However, the normality assumption for the coefficients might by considered a limitation of this model, since this assumption is not appropriate in all situations (Train, 2003, pp.97–114). For instance, the normality assumption for the price coefficient in a probit model with random taste variation implies that some respondents would have a positive price coefficient (i.e. positive preference for the price), as the normal distribution allows for negative as well as positive values. In this case, other distributions might be more appropriate, but this cannot be accommodated within the probit model yet. Other models can be specified for different research purposes (see Train, 2003, pp.17–18 for more details).

2.5.2 Model Identification

Discrete choice models depend on the structure of the random errors as well as the specification of the population mean utility function. However, there are two properties of discrete choice models that are implied by utility theory which could affect the specification of the utility function: only the difference in utility matters and the scale of utility is arbitrary. These properties affect the identifiability of the model parameters as described below.

- **Only Difference in Utility Matters**

  The choice probability of a particular alternative is determined by comparing its utility with the utilities for the offered alternatives in the same choice set, as shown in equation (2.5.1). This probability can be written as

$$P_{ijs} = P(U_{ijs} - U_{its} > 0 \quad \forall t \neq j). \tag{2.5.2}$$

This indicates that the choice probability of any alternative depends on the difference between the utilities only, and hence the absolute value of the utility is irrelevant to the choice behaviour. Therefore, adding a constant α to all alternatives' utilities does not change the alternative with the highest utility, and results in the same choice probability. as shown below.

$$\begin{aligned} P_{ijs} &= P\left[g(\boldsymbol{x}_{ijs}) + \alpha + \epsilon_{ijs} > g(\boldsymbol{x}_{its}) + \alpha + \epsilon_{its} \quad \forall t \neq j\right], \\ &= P\left[g(\boldsymbol{x}_{ijs}) + \epsilon_{ijs} > g(\boldsymbol{x}_{its}) + \epsilon_{its} \quad \forall t \neq j\right]. \end{aligned} \tag{2.5.3}$$

This makes it impossible to estimate a unique value for α, since any value gives the same choice probability. To make such constant identifiable, the constant of one alternative should be normalised to zero and estimate the other constants relative to the normalised one (Train, 2003, pp.19-23). To produce utility value within the required QALY scale, we set $\alpha = 1$ corresponds to the utility value of the full health state that is represented by the best possible health state defined by the classification system.

- **The Scale of Utility is Arbitrary**

  Multiplying the utility by any positive constant does not change the alternative with the highest utility; the scale of utility is irrelevant to the choice behaviour. Hence, the two models defined below are equivalent.

$$\begin{aligned} U^1_{ijs} &= g(\boldsymbol{x}_{ijs}) + \epsilon_{ijs}, \\ U^2_{ijs} &= \lambda g(\boldsymbol{x}_{ijs}) + \lambda\epsilon_{ijs} \quad \forall \lambda > 0. \end{aligned} \tag{2.5.4}$$

The probability of choosing alternative $\boldsymbol{x}_j$ from a set of alternative $j = 1, \ldots, J$ on the transformed scaled becomes

$$\begin{aligned} P_{ijs} &= P\left[\lambda g(\boldsymbol{x}_{ijs}) + \lambda\epsilon_{ijs} > \lambda g(\boldsymbol{x}_{its}) + \lambda\epsilon_{its} \quad \forall t \neq j\right], \\ &= P\left[g(\boldsymbol{x}_{ijs}) + \epsilon_{ijs} > g(\boldsymbol{x}_{its}) + \epsilon_{its} \quad \forall t \neq j\right], \end{aligned} \tag{2.5.5}$$

which is the same as the choice probability on the non-transformed scaled in

equation (2.5.1). Again, it is not possible to identify a unique value of λ, as different values give the same results for the choice probability. Therefore, the scale of utility should be normalised in order to be able to estimate unique utility values. Normalising the scale of utility is equivalent to normalising the variance of the random component of the utility (ϵ_{ijs}). This follows since the scale of utility is usually defined by the variance of ϵ_{ijs} and this variance increases by λ^2 when multiplying the utility by λ. Normalising the utility scale varies depending on the property of the variance of the random errors. Here, we briefly discuss the normalising issue for independent and identical errors, heteroscedastic errors and correlated errors, as described by Train (2003, pp.23–29).

1. **Independent Errors**

   Assume that the random errors of the latent utility in equation 2.3.1 are independent and identically distributed. Then normalising the utility scale is equivalent to normalise any of those error variances to a specific number, since errors are identically distributed (all have the same variance).

   Typically, the error variance is normalised to 1. This is done by setting $\lambda = \frac{1}{\sigma}$ in equation (2.5.4), and the utility becomes equivalent to

   $$\frac{U_{ijs}}{\sigma} = \frac{g(\boldsymbol{x}_{ijs})}{\sigma} + \frac{\epsilon_{ijs}}{\sigma}, \quad j = 1, \ldots, J \tag{2.5.6}$$

   where the error variance becomes equal to 1, $Var(\frac{\epsilon_{ijs}}{\sigma}) = \frac{1}{\sigma^2}\sigma^2 = 1$.

   In this thesis, we scale the utility by setting the utility of death equal to zero, i.e. $\lambda = \frac{-1}{\beta_d}$ where β_d is the death coefficient as described in Section 2.3.3. Therefore, the error variance become equal to $\frac{\sigma^2}{\beta_d^2}$.

2. **Heteroscedastic Errors**

   In some cases, the variance of the random errors varies across different groups of the population. In such situations, the scale of utility is normalised by normalising the variance for one group and then estimating the variance for the other relative to that group.

3. **Correlated Errors**

   If the random errors are correlated over alternatives, the utility scale cannot be normalised by normalising the variance of one alternative. In this situation, it is more appropriate to consider the variance of utility differences. Normalising one variance of the error differences leads to set the scale of the utility differences, and consequently the scale of the utility.

2.5.3 Multinomial Logit Model

In this section, we consider the case where the random errors in the utility model are independent and identically distributed with a type 1 extreme value distribution, to model DCE data using the multinomial logit model (MNL) model. In general, if a random variable X follows a type 1 extreme value distribution, the probability density function (pdf) is defined as

$$f_X(x) = \frac{1}{\sigma} \exp\left(\frac{-x+\mu}{\sigma}\right) \exp\left[-\exp\left(\frac{-x+\mu}{\sigma}\right)\right], \quad -\infty < x < \infty \tag{2.5.7}$$

where μ and σ are the location and the scale parameters, respectively. The random variable X has mean $E(X) = \mu + 0.5722\sigma$ and variance $Var(X) = \frac{1}{6}\pi^2\sigma^2$. The mean of the random error defined in Equation (2.3.1) is required to be zero in order to interpret the function $g(\boldsymbol{x}_{ij})$ as the population mean utility. Therefore, we set the location parameter $\mu = -0.5722\sigma$.

The cumulative distribution function (cdf) is given by

$$F_X(x) = \exp\left[-\exp\left(\frac{-x+\mu}{\sigma}\right)\right]. \tag{2.5.8}$$

Now assuming that the utility is defined as in equation (2.3.1), then the choice probability that individual i chooses alternative $\boldsymbol{x}_j$ from a set of J possible alternatives

in a choice set can be written as

$$P_{ijs} = P\left[\epsilon_{its} < g(\boldsymbol{x}_{ijs}) + \epsilon_{ijs} - g(\boldsymbol{x}_{its}) \quad \forall t \neq j\right]. \tag{2.5.9}$$

By conditioning on ϵ_{ijs} the choice probability can be computed as

$$P_{ijs} = \int_{-\infty}^{\infty} \prod_{t \neq j} P(\epsilon_{its} < g(\boldsymbol{x}_{ijs}) + \epsilon_{ijs} - g(\boldsymbol{x}_{its}) | \epsilon_{ijs}) f(\epsilon_{ij}) d\epsilon_{ij}, \tag{2.5.10}$$

where

$$f(\epsilon_{ijs}) = \frac{1}{\sigma} \exp\left(\frac{-\epsilon_{ijs} + \mu}{\sigma}\right) \exp\left[-\exp\left(\frac{-\epsilon_{ijs} + \mu}{\sigma}\right)\right],$$

and the conditional choice probability is

$$P(\epsilon_{its} < g(\boldsymbol{x}_{ijs}) + \epsilon_{ijs} - g(\boldsymbol{x}_{its}) | \epsilon_{ijs}) = \exp\left[-\exp\left(\frac{g(\boldsymbol{x}_{its}) - g(\boldsymbol{x}_{ijs}) - \epsilon_{ijs} + \mu}{\sigma}\right)\right]$$

Therefore, the choice probability of choosing alternative $\boldsymbol{x}_j$ is derived as

$$\begin{aligned}
P_{ijs} &= \int_{-\infty}^{\infty} \prod_{t \neq j} \exp\left[-\exp\left(\frac{g(\boldsymbol{x}_{its}) - g(\boldsymbol{x}_{ijs}) - \epsilon_{ijs} + \mu}{\sigma}\right)\right] \\
&\quad \times \frac{1}{\sigma} \exp\left(\frac{-\epsilon_{ijs} + \mu}{\sigma}\right) \exp\left[-\exp\left(\frac{-\epsilon_{ijs} + \mu}{\sigma}\right)\right] d\epsilon_{ijs}, \\
&= \int_{-\infty}^{\infty} \prod_{t=1}^{J} \exp\left[-\exp\left(\frac{g(\boldsymbol{x}_{its}) - g(\boldsymbol{x}_{ijs}) - \epsilon_{ijs} + \mu}{\sigma}\right)\right] \times \frac{1}{\sigma} \exp\left(\frac{-\epsilon_{ijs} + \mu}{\sigma}\right) d\epsilon_{ijs}, \\
&= \int_{-\infty}^{\infty} \exp\left[-\exp\left(\frac{-\epsilon_{ijs} + \mu}{\sigma}\right) \sum_{t=1}^{J} \exp\left(\frac{g(\boldsymbol{x}_{its}) - g(\boldsymbol{x}_{ijs})}{\sigma}\right)\right] \\
&\quad \times \frac{1}{\sigma} \exp\left(\frac{-\epsilon_{ijs} + \mu}{\sigma}\right) d\epsilon_{ijs}.
\end{aligned}$$

The computation of this integral can be simplified by defining $z = \exp\left(\frac{-\epsilon_{ijs} + \mu}{\sigma}\right)$ where

$dz = -\frac{1}{\sigma}\exp\left(\frac{-\epsilon_{ijs}+\mu}{\sigma}\right)d\epsilon_{ijs}$, and $c = \sum_{t=1}^{J}\exp\left(\frac{g(\boldsymbol{x}_{its})-g(\boldsymbol{x}_{ijs})}{\sigma}\right)$. Thus, it becomes

$$\begin{aligned} P_{ijs} &= \int_{\infty}^{0} -\exp(-cz)dz \\ &= \frac{1}{c}\exp(-cz)|_{\infty}^{0} \qquad (2.5.11) \\ &= \frac{1}{c} \\ &= \frac{\exp\left(\frac{g(\boldsymbol{x}_{ijs})}{\sigma}\right)}{\sum_{t=1}^{J}\exp\left(\frac{g(\boldsymbol{x}_{its})}{\sigma}\right)}. \qquad (2.5.12) \end{aligned}$$

The MNL model is then given by equation (2.5.12).

This model has been used in many choice studies due to its simplicity and having a closed form for the choice probability and consequently the likelihood function. This property of the MNL model simplifies the calculation of many statistical quantities and optimality criteria that are used to construct choice designs, as will be seen in the following chapters when generating choice design.

This thesis considers including death state in the choice design for scaling purpose. Therefore, it is important to derive the formula of the MNL model when one state in a choice set is death. Under RUT and assuming that in a choice set that includes death $\boldsymbol{x}_{iJs} = \boldsymbol{x}_{ids}$, i.e. $C_s = \{\boldsymbol{x}_{i1s}, \ldots, \boldsymbol{x}_{i(J-1)s}, \boldsymbol{x}_{ids}\}$, then individual i will choose alternative $\boldsymbol{x}_j$ for all $j \neq J$ from such choice set only if $U_{ijs} > U_{its}$ for all $t \neq j$ where $U_{ijs} > U_{ids} = 0$ means that $\epsilon_{ijs} > -g(\boldsymbol{x}_{ijs})$.

Thus, conditioning on $\epsilon_{ijs} > -g(\boldsymbol{x}_{ijs})$, the choice probability of alternative $\boldsymbol{x}_{ijs}$ in equation (2.5.10) can be written as

$$\begin{aligned} p_{ijs} &= \int_{-\infty}^{\infty}\prod_{t\neq j} P(\epsilon_{its} < g(\boldsymbol{x}_{ijs}) + \epsilon_{ijs} - g(\boldsymbol{x}_{its})|\epsilon_{ijs} > -g(\boldsymbol{x}_{ijs}))f(\epsilon_{ijs})d\epsilon_{ijs}, \\ &= \int_{-\infty}^{\infty}\prod_{t\neq j} P(\epsilon_{its} < g(\boldsymbol{x}_{ijs}) + \epsilon_{ijs} - g(\boldsymbol{x}_{its})|\epsilon_{ijs})I(\epsilon_{ijs} > -g(\boldsymbol{x}_{ijs}))f(\epsilon_{ijs})d\epsilon_{ijs}, \end{aligned}$$

where $I(\epsilon_{ijs} > -g(\boldsymbol{x}_{ijs})) = 1$ if $\epsilon_{ijs} > -g(\boldsymbol{x}_{ijs})$ and 0 otherwise. Thus, the choice probability can be computed as

$$P_{ijs} = \int_{-g(\boldsymbol{x}_{ijs})}^{\infty} \prod_{t \neq j} P(\epsilon_{its} < g(\boldsymbol{x}_{ijs}) + \epsilon_{ijs} - g(\boldsymbol{x}_{its}) | \epsilon_{ijs}) f(\epsilon_{ijs}) d\epsilon_{ijs}. \tag{2.5.13}$$

It can be shown from equation (2.5.11) that the choice probability is equal to

$$P_{ijs} = \frac{\exp\left(\frac{g(\boldsymbol{x}_{ijs})}{\sigma}\right)}{\sum_{t=1}^{J-1} \exp\left(\frac{g(\boldsymbol{x}_{its})}{\sigma}\right)} \left\{1 - \exp\left[-\sum_{t=1}^{J-1} \exp\left(\frac{g(\boldsymbol{x}_{its}) + \mu}{\sigma}\right)\right]\right\}, \tag{2.5.14}$$

where P_{ijs} is the choice probability of health state compared to death. Hence the choice probability of the death state can be written as

$$\begin{aligned} P_{iJs} &= 1 - \sum_{j=1}^{J-1} P_{ijs}, \\ &= \exp\left[-\sum_{t=1}^{J-1} \exp\left(\frac{g(\boldsymbol{x}_{its}) + \mu}{\sigma}\right)\right]. \end{aligned} \tag{2.5.15}$$

Inference about the utility function can be carried out by analysing choices observed for each alternative in a choice set, since choices depends on the mean utility values, $g(\boldsymbol{x}_{ijs})$, that is a function of the parameters $\boldsymbol{\beta}$. The choices from a sample of N respondents made over all alternatives in S choice sets can be considered as independent draws from a multinomial distribution, due to the assumption of independent random errors over alternatives and respondents. Therefore, the likelihood function of the multinomial logit model for the observed choices over S choice sets, can be written as

$$L(\mathbf{y}|\boldsymbol{\beta}, \sigma) = \prod_{s=1}^{S} \prod_{j=1}^{J} \prod_{i=1}^{N} p_{ijs}^{y_{ijs}}, \tag{2.5.16}$$

where y_{ijs} is defined as

$$y_{ijs} = \begin{cases} 1 & \text{if health state } \boldsymbol{x}_{ijs} \text{ is chosen,} \\ 0 & \text{otherwise.} \end{cases}$$

and P_{ijs} is the corresponding probability for health state $\boldsymbol{x}_{ijs}$ which is defined as in equation 2.5.12 if the underlying choice set does not include death state, and as in equation 2.5.14 and 2.5.15 for health state in a death choice set and for death state, respectively.

The log likelihood function is then defined as

$$l(\mathbf{y}|\boldsymbol{\beta}, \sigma) = \sum_{s=1}^{S} \sum_{j=1}^{J} \sum_{i=1}^{N} y_{ijs} \log P_{ijs}. \tag{2.5.17}$$

The full likelihood for our analysis is given in Chapter 4.

2.6 Summary

In this chapter, we have reviewed the use of the QALY in health economics as a measure for health in terms of quality and quantity of life. To compute the QALY gained of a health intervention, health outcomes need to be described and then evaluated using a preference method to elicit health state utility values. This chapter introduced different classification systems to describe health outcomes, which can be generic (e.g. EQ-5D) or specific (e.g. AQL-5D). It then reviewed various valuation techniques to derive the health state utilities such as TTO and SG methods. These methods have been criticized by many researchers due to their complexity. Therefore, there has been

increased interest in using discrete choice experiments (DCEs) as an alternative to those methods. A description of DCEs and modelling their choice data, particulary using the MNL model, have been presented in this chapter. However, DCE methods still need plenty of work before they can be established as an alternative for the ordinal methods such as TTO and SG methods, particularly in terms of choice design construction and selecting choices presented to respondents such that reliable estimates for health state utilities are obtained, as will be discussed in Chapter 3.

Chapter 3

Literature Review: DCEs and their Design in Health Economics

3.1 Introduction

Discrete choice experiments (DCEs) have been widely used in many market research studies (e.g., transport, environment, banking etc.) to identify the most important features/attributes in specific products or services from consumers perspective. In the last 15 years, there has been increasing interest in using DCEs in a health economics evaluation context. In particular, recently, this method has been increasingly used to estimate health state utilities/quality weights within the QALY framework. This technique, as discussed in Chapter 2, seems to be a promising alternative to cardinal methods for eliciting utility values within the required QALY scale. Nevertheless, it requires considerable developments, particularly in terms of designing the choice exper-

iment (e.g., selecting alternatives, constructing the choice tasks, number of choices) to improve the quality of the choice data and, hence, provide a more reliable assessment for health state utilities.

There is increasing evidence that using more sophisticated methods to generate DCE designs, together with appropriate analytical techniques, improves the quality of the choice data collected and the final outputs (de Bekker-Grob et al., 2010). There has been development in the methodology of deriving efficient DCEs in various areas outside health economics. In this chapter, therefore, we recognise how health economics has benefited from these improvements in DCE designs, and identify the remaining design issues before a robust methodology for generating DCEs for valuing health utilities can be developed.

The chapter begins with reviewing the applications of DCEs in health economics, specifically the experimental designs used to construct the choice designs for valuing utilities and their methodological issues in Section 3.2. In Section 3.3, we discuss the main experimental design considerations in a health valuation context, and the need for an algorithmic procedure to search for an optimal, or near optimal, design that satisfies the study constraints. Section 3.4 reviews the main concepts of optimal design theory and the statistical measures known as optimality criteria. Optimal design theory has been used for linear and non-linear models; however, since our project focuses on generating DCEs for discrete choice models (DCMs), we will place greater emphasis on reviewing optimal design for nonlinear models, particularly choice models. In Section **??**, we present the main problem with generating optimal choice designs for DCMs, and then discuss the use of a Bayesian approach to overcome this problem. We also review different methods for deriving Bayesian design criteria in Section 3.6. In the last section, we review Bayesian design strategies and algorithms used in the design literature to construct choice designs for DCMs, which then might be applied in health economics to improve the construction of choice designs for health valuation.

3.2 DCEs in Health Economics

The use of DCEs to value health benefits has increased rapidly since they were introduced to health economics in 1990 by Propper. The DCE techniques, also known as conjoint analysis (CA), have been developed in health economics to simplify the valuation process for respondents, and assess health benefits beyond direct health.

We briefly summarise the applications of DCEs in health economics, with greater emphasis on the applications that are similar to our design objective, that is, constructing choice design for valuing health state utilities within the QALY framework, in Section 3.2.1. Then, we discuss experimental design methods used in those applications to construct choice designs together with their methodological issues in Sections 3.2.2 and 3.2.3, respectively.

3.2.1 DCEs in Health Economics: A Review

The DCE technique was first used in health economics to value a health service using patient preference in Propper (1990). Since then, many discrete choice studies have been used to assess the benefits of different health services beyond simple health outcomes. For example, Ryan and Farrar (1994) used conjoint analysis to investigate the trade-offs that patients make between the location of the clinics and waiting times in the provision of orthodontic services. Their study shows the importance of these non-health outcomes, as patients were indeed concerned about these features of the overall health service provided. There are other similar studies, cited in Ryan (1999), that looked at valuing the benefits of different health services beyond direct health outcomes using patient preferences, and the trade-off made between direct health outcomes and

non-health outcomes (e.g., Ryan and Hughes, 1997; Bryan et al., 1998; and Pol and Cairns, 1998).

Over the last two decades there have been a considerable number of applications of DCEs in health economics that cannot be discussed in detail here. So it is worth mentioning several systematic reviews conducted in the health economic literature. The first substantial review, by Ryan and Gerard (2003), identifies 34 applications of DCEs conducted between 1990 and 2000, the majority of them (about 58%) in the UK. This baseline review (covering the period 1990–2000) was updated by two other systematic reviews: de Bekker-Grob et al. (2010), who identify 114 applications of DCEs between 2001 and 2008; and Clark et al. (2014), who update the latest review up to 2012. Both later reviews follow the same search procedure as the baseline review; thus they use the same database to find the related DCE applications (PubMed) and the same search criteria and terminologies.

The latest review shows a dramatic increase in use of DCEs in health economics, with 179 applications published in four years. That is on average 45 applications per year compared with just 14 applications per year during the period 2001–2008. The reviews show that while the UK remain the largest producer of these applications among the developed countries, the number of applications from lower income countries such as Kenya, China and Thailand increased.

The majority of these applications consider valuing health outcomes beyond direct health and trading between direct health outcomes and process attributes, with recent reviews showing more interest in valuing health outcomes within a QALY framework. Focusing on those applications which used DCEs to value health outcomes for provision of the QALY weights considered in these reviews, we found five key studies, as shown in the first five lines of Table 3.1.

During the period 1990–2000, there were no applications that aimed to estimate utility weights within the QALY framework. In the period 2001-2008, two studies (2%

Table 3.1: Method to create choice set for DCE applications conducted to value health state utilities in health economics together with algorithms and software used to generate the choice design

Study	***Design Type***	***Method to Create Choice Set***	***Software***	***Anchoring Utility Values***
2001-2008:				
1.Ryan et al. (2006)	Orthogonal design	Computer-based software	SPPED	Value of best and worst health states
2.Burr et al. (2007)	Orthogonal Design	Foldover technique	SAS	Value of best and worst health states
2009-2012:				
1.Brazier et al. (2009)	$\mathcal{D}$-optimum design	Huber and Zwerina (1996)	SAS	Including death state
2.Stolk et al. (2010)	Bayesian $\mathcal{D}$-optimum design	Random generation	Excel	1- TTO value of worst health state 2- Including death state
3.Bansback et al. (2010)	$\mathcal{D}$-optimum design	Kuhfeld (2000)	SAS	Including years of survival as a separate attribute
2013-2014:				
1.Gu et al. (2013)	$\mathcal{D}$-optimum design	Huber and Zwerina (1996)	SAS	Value of best and worst health state
2.Prosser et al. (2013)	Full factorial design	Not reported	Not reported	Including years of survival
3.Ramos-Goñi et al. (2013)	Bayesian $\mathcal{D}$-optimum design	Random generation	Not reported	1- TTO value of worst health state 2-Including death state
4.Viney et al. (2014)	Orthogonal design	Foldover technique	SAS	1-Including years of survival 2-Including death state
5.Bansback et al. (2014)	$\mathcal{D}$-optimum design	Zwerina et al. (1996)	SAS	Including years of survival

of the applications) had this as their main objective. In particular, two studies use DCEs as an alternative to cardinal methods such as SG and TTO techniques (Ryan et al., 2006; Burr et al., 2007). The amount of work in this area has continued to rise: during the period 2009–2012 Clark et al. found four studies directed towards QALY estimates, of which three are of particular interest. These analyses used DCEs to estimate health state utility for the AQL-5D and the EQ-5D states as in Brazier et al. (2009) and (Stolk et al., 2010), respectively, and constructed a choice experiment to resemble the TTO task by including years of survival as an attribute in the choice experiment (Bansback et al., 2010).

Using the same search terms as in the previous systematic reviews, we found that the number of applications of DCEs in the health economics field has continued to increase rapidly. Our current search shows that there have been around 200 applications of DCEs applied to valuing health outcomes in different areas of health economics from 2013 onwards. In addition, the current search shows an increase in the number of analyses that used DCE to estimate health state utility within the QALY scale, in which we found five studies published in 2013 and 2014. Table 3.1 summarises these applications together with the previous findings, showing also the methods used to create choice designs and anchor health state utility values onto the required QALY 0-1 scale.

3.2.2 Methods Used to Create Choice Sets

An important phase in constructing any DCE is the choice of experimental design, i.e. how to combine the attribute levels to create choices and then group them efficiently into a choice set. The experimental design is used to reduce the number of possible choices to manageable numbers while ensuring the analyst is able to estimate the main effects of interest and possible interactions. The majority of the relevant studies – all the studies shown in Table 3.1 – used a fractional factorial design approach, except Prosser et al. (2013) who used full factorial design that includes all possible combinations of

the attributes' levels in the choice designs.

In general, the studies show that the methods employed to construct the choice design during the last decade have all been similar. Orthogonal designs, obtained from orthogonal arrays available in different statistical software such as SAS, have been used in some studies, such as Ryan et al. (2006), Burr et al. (2007) and Viney et al. (2014), to construct the fractional factorial design. These arrays have the properties of orthogonality (each pair of levels of different attributes appears an equal number of times in the alternatives) and level balance (levels of attributes appear an equal number of times, i.e. for an attribute with five levels each level has to occur at exactly one-fifth of the design points). For binary DCEs (i.e. would you choose a specific health condition, yes or no?) the profiles generated from the orthogonal design are the choices. However, for multiple DCEs, choosing between two or more profiles, a method is needed to move from an orthogonal design to the choice sets. Louviere et al. (2000) proposed two methods to create choice sets from orthogonal designs:

- 'Foldover', also known as the shifting and cycling method, begins by constructing a mirror image of the original design in which each profile in the orthogonal design is paired with its foldover or complement. Thus for five-level attributes we replace $0 \rightarrow 1$, $1 \rightarrow 2$, $2 \rightarrow 3$ and $4 \rightarrow 0$ for five levels, for example, AQL-5D state 11233 would be paired with 22344. Each pair of the original profile and its complement then creates a single choice set; hence all the combinations of all pairs construct the choice design.

- 'Foldover with random pairing', where profiles in the original orthogonal design and their foldover are randomly paired. That is, all profiles and their foldovers are used in the choice sets, but a profile from the original design and its mirror image are randomly paired, not used in the complementary pairs.

In recent marketing design literature, there has been a development of using optimal design theory in the experimental design used to construct choice experiments. The

optimal designs are constructed based on different statistical efficiency measures called 'optimality criteria', as will be discussed in Section 3.4.1. The $\mathcal{D}$-optimality criterion, which is related to minimising the determinant of the variance-covariance matrix of the parameter estimators, is one of the most commonly used efficiency measures. This is presumably due to its simplicity, general applicability and lower computational cost compared with other criteria.

Health economists show more awareness of the importance of efficient designs and increased interest in using computerised search algorithms (Huber and Zwerina, 1996; Zwerina et al., 1996 and Kuhfeld, 2000) to generate choice sets. The algorithms search for the choice design that minimises the $\mathcal{D}$-optimality criterion given prior estimates for the unknown model parameters. Prior information is required in this case, since the optimality criterion for a nonlinear model, unlike that for a linear model, depends on the values of the unknown parameters, as will be illustrated in Section **??**. Nevertheless, designs constructed for valuing utilities using these search algorithms (e.g., Brazier et al., 2009; and Bansback et al., 2010 and 2014; Gu et al., 2013) were mainly constructed ignoring the dependency of the choice design on the parameters by assuming zero priors for the preference parameters.

A recent development is to use what is called Bayesian optimal designs that incorporate genuine prior assumptions about the unknown parameters in generating the choice design to improve the statistical efficiency (e.g., Sándor and Wedel, 2001, 2002, 2005; Rose and Bliemer, 2008; Kessels et al., 2009; and others that will be discussed further in Section 3.7.2). Up to now, there have been only two studies, Stolk et al. (2010) and Ramos-Goñi et al. (2013), which attempt to use Bayesian design to construct the choice design for estimating health state utilities. Here Bayesian optimum design is generated, simply, based on random search algorithm that selects the optimal choice design using random search over a large number of choice designs. The algorithm uses Monte Carlo simulation to determine the design efficiency by computing the $\mathcal{D}$ criterion value (Bliemer et al., 2008). The algorithm returns the design with

the desirable number of choice sets that minimises the $\mathcal{D}$ criterion value over the prior distribution of the parameters.

None of these studies used advanced search algorithms to construct the Bayesian choice designs, and nor did they investigate the effect of the prior knowledge on improving the design properties and final model outputs (utility estimates). This area of research deserves more attention from health economics researchers, to produce efficient choice designs and improve the quality of data collected. This will be one of our main concerns throughout the thesis: to develop an efficient methodology for constructing choice design for estimating health state utility within the QALY framework.

Despite the fact that the current review finds some development in methods used to construct DCEs for valuing health state, there are still crucial methodological issues associated with the experimental designs and algorithms used to generate choice designs. We refer the reader to de Bekker-Grob et al. (2010) and Johnson et al. (2013) for a discussion of general issues related to using DCEs in health evaluation studies, while, in the following section, we discuss those issues related specifically to the design methods and search algorithms used to construct the DCEs for valuing health state utilities.

3.2.3 Methodological Issues

In this section, we discuss the methodological issues related to constructing choices design for valuating health state utilities. In particular, we discuss the issues with experimental designs and algorithms used so far in health economics to generate efficient choice design for provision utilities within the QALY scale. We stratify these issues by the type of experimental design used.

1. **Orthogonal Arrays**

   The construction of orthogonal choice designs is based on orthogonal arrays.

These arrays are used to construct the first option (here health state) in each choice set, and then to create the subsequent option(s) using the foldover method proposed in Louviere et al. (2000) or one of the pairing strategies suggested in Street et al. (2005). These orthogonal arrays have desirable statistical properties that allow for estimating the main effects of interest independently, while assuming all interactions are zero. However, they might not be appropriate for constructing choice designs. This is because these orthogonal arrays are based on linear design principles and using such designs to construct DCEs ignores the nonlinear nature of the choice models, which reduces the efficiency of the choice designs (Ryan et al., 2006).

Another issue is that the efficiency of orthogonal design is optimised based on the fact that choices are made randomly. This is true only under the restrictive assumption that all the preference parameters identified in the utility model are equal to zero. Thus, if the true parameter values deviate from zero, then the orthogonal design will not be optimal for estimating the true parameter effects accurately (Stolk et al., 2010).

In addition, orthogonal arrays select health states irrespective of the realism of their attribute level combinations. This might result in implausible/unrealistic health states that increase the variability in the responses (Johnson et al., 2013). Also, the subsequent pairing method, such as foldover, may produce 'dominant' choices (e.g., it might pair AQL-5D health state 11123 with 22234, where 11123 is always preferred). This type of choice reduces the efficiency of the designs, as will be discussed in Section 3.3.

2. **Locally $\mathcal{D}$-efficient Design**

A key determinant of the efficiency of a choice design is the algorithm used to generate the choice sets. For example, Huber and Zwerina (1996) identify four design principles they believe to be characteristic of optimal choice designs. These are: orthogonality (each pair of levels of different attributes appears an

equal number of times in the alternatives) and level balance (levels of attributes appear an equal number of times); minimum overlap (attribute levels are not repeated within a choice set); and utility balance (choices within a choice set are equally attractive). Zwerina et al. (1996) state that for certain families of shifted designs (e.g. designs constructed using the orthogonal array and foldover method) and assuming that all the preference parameters are zero, these designs will jointly satisfy the four principles and hence are optimal. This has not been proved analytically, but Street and Burgess (2007) show that these designs are more efficient than other designs that are not generated from those principles.

However, it may not be possible to construct a choice design that satisfies all the four principles because, for particular design problems, these principles might conflict with each others (Huber and Zwerina, 1996). In addition, Street and Burgess (2007) argue that even satisfying these four principles does not guarantee that the resulting design is optimal, nor that it can estimate all the main effects of interest. Also, the orthogonality property might lead to less efficient designs, as we discussed earlier.

Other studies, such as Bansback et al. (2010, 2014), used a computerised search algorithm (see Zwerina et al., 1996; or Kuhfeld, 2000) to calculate the $\mathcal{D}$-optimum design instead of using these principles. These algorithms, described in Section 3.7.1, optimise the design efficiency using particular exchange procedures in which the efficiency of the design is determined by directly computing the $\mathcal{D}$ criterion value given the prior estimates. The algorithms return the choice design with the required number of choice sets that minimises (within the limit of the search) the $\mathcal{D}$ criterion value. However, they require a candidate set that includes all or some of possible health states defined by a classification system, which is built using full or fractional factorial design and typically has orthogonal and level balance properties. Restricting the elements of the candidate set to those properties (i.e. orthogonality and level balance) might produce implausible health state as in this case, once again, health states are selected irrespective to their attribute levels

combinations as discussed earlier in the orthogonal designs. This reduces the efficiency of the choice design.

In addition, to construct a choice design using these algorithms, prior estimates are usually required for the unknown model parameters, because of the nonlinear nature of DCEs. Nevertheless, efficient designs constructed using these search algorithms are usually derived based on zero priors for the preference parameters. This approach assumes no preference for any attribute levels, and hence choices are made randomly. However, this is unrealistic, since respondents are assumed to make their choices by trading-off between attribute levels of the alternatives provided. Thus, a design based on this assumption might not be optimal and might produce insufficient information to estimate true parameter effects. Huber and Zwerina (1996) illustrated that incorrectly assuming zero parameter values might require from 10% to 50% more respondents to perform the experiment compared with a design built based on more realistic prior estimates.

Also, for choice design with ordinal attributes, this assumption might result in more dominant choices, i.e. choices always preferred by all respondents as they have the best levels in all attributes as will be described in Section 3.3, than a design based on more reasonable point parameter estimates or even more informative prior distributions. This is because incorporating prior information about the preference parameters, as opposed to zero priors, can account for the differences between the attribute levels of the alternatives presented in a choice task and, hence, the mean utility differences between alternatives. Thus, the assumption of zero priors can result in alternatives with high level differences which decrease the efficiency of the design, i.e. the amount of information obtained from the design to infer the preference parameters associated with the attribute levels of the underlying classification system, (Kessels et al., 2011b), where less efficiency design will result in larger standard errors of the estimated parameters and, hence, mean utilities.

3. **Bayesian $\mathcal{D}$-efficient Design**

   The Bayesian designs in Stolk et al. (2010) and Ramos-Goñi et al. (2013) incorporate more realistic prior information through the prior distribution; however, they were constructed based on a random search algorithm. The algorithm determines design efficiency, by computing the $\mathcal{D}$ criterion value over the prior distribution, for many designs randomly selected from the full factorial design, and then returns that design with minimum criterion value. The procedure often produces dominant and implausible choices in the final design, which therefore has to be examined manually for the presence of implausible and dominant choices.

   Therefore it would be better to use more advanced and flexible experimental design algorithms that allow for the incorporation of prior information in the design and control the presence of the dominant and implausible health states, instead of manually having to check for the plausibility of the attribute levels combinations generated by the simpler algorithms.

   There are more sophisticated search algorithms developed in marketing especially for generating Bayesian choice designs that might be able to do so, but these have so far received little attention from health economists. These seek to improve the search procedure and hence the efficiency of the choice design. Examples are Kessels et al. 2009, 2010, 2011; and Rose and Bliemer 2012 (see Section 3.7.2 for more details). In this thesis, therefore, we will mainly consider the issue of improving experimental choice designs used to estimate health state utilities within the QALY framework by using these latest developments in Bayesian optimal design and employing and adapting their design algorithms to our design problem.

Besides the experimental design issues discussed above, there are other non-mathematical issues, such as the complexity of the choice task, the number of choice sets that respondents could answer and other psychological issues, which are beyond the scope of this review. We refer the reader to Street and Burgess (2007, pp.11–13)

who provided a good reference for these issues. In the following section, we discuss the main design considerations that should be addressed when constructing choice designs for health evaluation, in general, and for health state utility estimation, in particular.

3.3 Context-specific Design Considerations

Besides the general methodological issues illustrated earlier, there are some other aspects that should be considered when constructing choice sets for health evaluation studies.

- **Constraint on Implausible Alternatives**

  In DCEs, choices are collected based on hypothetical profiles or alternatives defined by the attribute levels of the underlying classification system as illustrated in Section 2.3. However, some of the attribute level combinations are unrealistic or illogical and would make no sense to respondents. For example, consider the AQL-5D classification system introduced in Table 2.1, where each health state is defined by five-digits each reflecting the degree of severity (ordered from best level 0, to worst level 4) of particular attributes. In this system AQL-5D health state 00140 does not make a good sense, since in a situation where a person never has a good night's sleep (level 4 for sleep attribute), it would not be logical to assume that he/she would have no limitation at all with his/her daily activities (level 0 for the activity attribute).

  Thus, for any AQL-5D health state $\boldsymbol{x}_{ij}$ that is represented by a vector of dummy

variables with elements defined as

$$x_{\lambda\delta} = \begin{cases} 1 & \text{if attribute } \delta \text{ of health state } \boldsymbol{x}_{ij} \text{ is at level } \lambda \text{ or higher,} \\ 0 & \text{otherwise,} \end{cases}$$

if two of the first three attributes are at very mild level, say 0, while the rest of the attributes are at severe level, say 3 or 4, for example AQL-5D health state 00044, then the health state is likely to be unrealistic or very rare. Formally, the AQL-5D health state can be defined as unrealistic if

$$\sum_{\lambda=1}^{4}\sum_{\delta=1}^{3} x_{\lambda\delta} < 2 \quad \& \quad \sum_{\lambda=1}^{4}\sum_{\delta=4}^{5} x_{\lambda\delta} \geqslant 4. \tag{3.3.1}$$

Respondents might find it difficult to evaluate such alternatives, since they do not make good sense to them. This might reduce the response efficiency, and hence affect the precision and the accuracy of the parameter estimates which is typically the main interest of experimenters. Therefore, it is important to prevent this type of combination appearing in the choice designs.

- **Constraint on Dominant Alternatives**

  Dominant alternatives occur with ordinal attribute levels, as in health classification systems where attribute levels are ordered based on their severity from best to worst or vice versa. An alternative that has all or most of its attribute levels better than the other is said to be dominant.

  The alternative might dominate all the possible level combinations defined by a classification system, or only the alternatives offered in a particular choice set. To illustrate these situations, return to the AQL-5D example. The best health state (00000; no concern about having asthma, no feeling short of breath, no problem with weather and pollution, no sleep problems, no limitation in activities at all) dominates all other possible health states defined by the classification system. This is because this health state has all the best level combinations, so it is clear that all respondents will choose the best health state no matter what other health

states are offered to them. On the other hand, a choice set such as $\{00011, 01212\}$ has an alternative that dominates the other option offered in this given choice set – health state 00011 has attribute levels that are better (or as good as) than the corresponding level of the other alternative – but it does not dominate all other possible level combinations defined by the classification system, such as 00000, or perhaps even 11000. Mathematically, for any two AQL-5D health states $\boldsymbol{x}_{ij}$ and $\boldsymbol{x}_{it}$ where $t \neq j$, health state $\boldsymbol{x}_{ij}$ is formally said to be dominant over $\boldsymbol{x}_{it}$ if

$$\sum_{\lambda=1}^{4} x_{\lambda\delta}^{j} \leqslant x_{\lambda\delta}^{t}, \quad \text{for all} \quad \delta = 1, \ldots, 5. \tag{3.3.2}$$

Informally, the term is also applied if the inequality holds for 'most' attributes.

This type of choice question does not provide valuable information regarding the preferences for the attributes. This follows since all respondents will choose the dominant alternative. Since these choices are modelled to infer respondents' preference for the attributes, as illustrated in Section 2.5, a choice design with many dominant choice sets will reduce the precision of the parameter estimates. Therefore it is usually better to avoid this type of question in the actual choice design. Nevertheless, it may occasionally be sensible to use them to test respondents' understanding for the choice task and the definition of the attribute levels.

- **Anchoring Health State Utility**

  As mentioned earlier in Chapter 2, a major problem with using DCEs for estimating health state utility is that the resulting estimates are not directly anchored on the dead-full health scale required to compute QALYs. Therefore, in this thesis, we suggest following the idea of including the death state in the choice design to anchor the utility values using the death coefficient, as discussed in Section 2.4.4. This idea was originally proposed in Brazier et al. (2009) and has been applied by many researchers, such as Stolk et al. (2010) and Ramos-Goñi et al. (2013). In these studies, the death state was either added manually to the original choice design as an extra choice task or added as a common option to

all choice sets, to perform a ranking exercise and estimate the death coefficient. However, this might affect the efficiency of the choice design, since the original design is optimised based on a particular criterion that does not account for the death state.

Therefore, it would be better if we could use an approach that allows us to involve the death state in the construction of the choice design. This will be another of our concerns in generating a choice design for valuing health state utility.

It is essential to construct a choice design for health evaluation studies that takes into account all of these features if we wish to gather information on choices as efficiently as possible. Constraints imposed on the implausible and dominant alternatives can act in opposition to the general aim to maximise the information that can be gained through the design. This, in addition to the complexity of constructing the choice design for a nonlinear DCM model, means that 'standard' designs such as orthogonal designs are unsuitable. Tailored approaches must be sought to generate optimal choice designs for health evaluation studies.

3.4 Optimality Theory

Constructing the choice tasks of DCEs, i.e. selecting the combinations of the attribute levels and grouping them efficiently into choice sets, requires an efficient experimental design that yield the maximum amount of information at the least cost and amount of work. We mentioned in the previous section that standard designs such orthogonal designs are not practical in many circumstance as they cannot cover the experimental constraints specified by the researchers. Also, they might not exist for complicated

and large choice design problems (e.g. designs with larger numbers of attributes and attribute levels, number of choice sets etc.).

An algorithmic optimal design methodology is required in such cases to generate the choice designs. Optimal designs (also called optimum designs) are constructed based on different statistical measures called optimality criteria. The selection of the design criterion is based on the objective of the experiment, as discussed in Section 3.4.1. Generally, optimal designs cannot be obtained theoretically and are constructed using search algorithms that use a specific search procedure to return the required number of choice questions with the specified number of alternatives that optimises the specified criterion given any design constraints (see Section 3.7 for more details about optimal design algorithms). The resulting optimal design is not the best of all possible designs; it is just the best found within the limitations of the search strategy and only with respect to the particular criterion chosen. Therefore, some authors word optimal.

The main advantage of optimal designs is that they are more flexible than the standard designs, particularly orthogonal designs, as they allow researchers to construct a good choice design, providing the required statistical properties and taking into account the design constraints. In addition, they are available for any design problem with any required number of choice questions, alternatives, and attributes, and for any type of model. Also, they may reduce the cost of experimentation in comparison with non-optimal designs, such as the full factorial designs, which require many choice tasks of the experiment to obtain the same precision in the parameter estimates. However, the reliance of the optimal designs on the model being specified before observing the effects of the experimental factors is a drawback, since this model dependence means that the quality of the data collected from the design and the final conclusion derived from analysing such data is based on the correctness of the model specified for generating the optimal design.

The first optimal design was constructed by Smith (1918) who stated the first optimality criterion for a linear regression problem and obtained an optimal design

based on what is called the **G**-optimality criterion. Optimal experimental designs for regression models subsequently received more attention in the design literature (Atkinson et al., 2007). However, the main contribution to the development of this area is due to Kiefer (1959, 1961, 1974) and Kiefer and Wolfowitz (1959, 1960, 1965) cited in Chaloner (1984). In-depth information regarding optimal experimental designs for linear models in a non-Bayesian framework can be found in works by Federov (1972) and Silvey (1980) (Atkinson et al., 2007).

For the past two decades, interest in Bayesian experimental designs has increased, with much work carried out to construct Bayesian designs for linear and nonlinear models. More details can be found in the review paper by Chaloner and Verdinelli (1995). Also, optimal experimental designs for both linear and nonlinear models have been discussed in depth by Atkinson et al. (2007). In this chapter, we mainly discuss Bayesian optimal designs for nonlinear models, and even more specifically only for DCMs. However, before that, we review some optimal design criteria and general aspects of the optimal design theory in Sections 3.4.1 and 3.4.2, respectively, and then illustrate the calculation of the optimal design for the general linear model in Section 3.4.3, to illustrate the difference between linear and nonlinear optimal design problems.

3.4.1 Optimality Criteria

Optimal designs are based on different statistical measures called optimality criteria. These criteria are also known as alphabetic optimality criteria, since they are named by different letters of the alphabet. The optimality criterion is a single value or measure used to explain how good a design is. This value summarises and describes the properties and variability of the parameter estimators. Thus, an optimal experimental design allocates the support points $\boldsymbol{x}_j$, $j = 1, \ldots, J$, of a design ξ in the experimental design region $\mathcal{X}$ such that the optimality criterion value is optimised. The vector $\boldsymbol{x}_j$ defines the levels of the explanatory variables of the j^{th} design point, and these are used to define the design matrix, $\mathbf{X}$, and hence summarise the amount of information

provided by an experiment by the information matrix of a design $\mathbf{M}(\xi)$.

The optimality criteria are typically defined as a function of the information matrix or its eigenvalues of a design ξ, denoted as $\mathbf{M}(\xi)$. The importance of this matrix is its proportionality to the inverse of the variance–covariance matrix of the parameter estimators of the model, since that is usually used as a measure of a good estimation procedure. In this section we give a brief overview of the most frequently used design criteria in practice. These are:

- **$\mathcal{D}$-optimality**: This criterion was introduced by Wald (1943), and has been the most extensively studied criterion in Bayesian and non-Bayesian design literature (Rady et al., 2009). It seeks to maximise the determinant of the information matrix of the design, denoted by $|\mathbf{M}(\xi)|$, or equivalently minimise the determinant of its inverse (the variance–covariance matrix of the estimators). Formally this can be written as

$$\mathcal{D} = |\mathbf{M}^{-1}(\xi)|. \tag{3.4.1}$$

  This criterion is typically used when the interest is in estimating all parameters of interest, $\boldsymbol{\beta}$, in a particular model precisely, since the optimal design minimises the volume of the confidence ellipsoid of the estimated parameters. The k^{th} root, where k is the number of parameters of interest, of the determinant of the information matrix is typically used to standardise the statistical measure to result in a measure for which the dimension of the model is irrelevant (Atkinson et al., 2007). Thus, the criterion would be written as

$$\mathcal{D} = |\mathbf{M}^{-1}(\xi)|^{1/k}. \tag{3.4.2}$$

  Other criteria, such as $\mathcal{D}_A$ (also known as generalised $\mathcal{D}$-optimality) and $\mathcal{D}_S$, are extensions of $\mathcal{D}$-optimality. The former is used if the interest is in minimising the variance of r linear combinations of the parameter estimates $\mathbf{A}\boldsymbol{\beta}^T$, where $\mathbf{A}$ is a $r \times k$ matrix with rank $r < k$; the latter is suitable when the interest is in

estimating a subset of the parameters as precisely as possible.

If $r = 1$, i.e. the interest is in estimating a single linear combination of the parameter estimates $\mathbf{c}\boldsymbol{\beta}^T$ with minimum variance, then $\mathcal{D}_A$ is reduced to what is called **c**-optimality criterion. In this case, the design criterion to be minimised is defined proportional to the variance of $\mathbf{c}\boldsymbol{\beta}^T$ as

$$var(\mathbf{c}\boldsymbol{\beta}^T) \propto \mathbf{c}\mathbf{M}^{-1}(\xi)\mathbf{c}^T, \tag{3.4.3}$$

where **c** is a row vector with dimension equal to the number of parameters k, and $\sum_{i=1}^{k} c_i^2 > 0$. We refer the reader to Atkinson et al. (2007, pp.135–150) for more details about these criteria.

- **G-optimality**: This criterion was named by Kiefer and Wolfowitz (1959) after it was first introduced by Smith (1918) who constructed a **G** optimal design for a regression model. This criterion looks for designs minimising the maximal prediction variance, $d_{max}(\xi)$, over the design region $\mathcal{X}$

$$d_{max}(\xi) = \max_{\boldsymbol{x}\in\mathcal{X}} d(\boldsymbol{x}, \xi), \tag{3.4.4}$$

where $d(\boldsymbol{x}, \xi) = f(\boldsymbol{x})\mathbf{M}^{-1}(\xi)f(\boldsymbol{x})^T$ is the standardised variance of the prediction at the design point $\boldsymbol{x}$, and $f(\boldsymbol{x})$ is a row vector of known functions of $\boldsymbol{x}$. Thus, a **G**-optimum design is calculated as the one that minimises $d_{max}(\xi)$. The computational time of this criterion grows exponentially as the design region $\mathcal{X}$ increases. Thus, to make this design criterion more feasible it is much more convenient to restrict the design region.

- **$\mathcal{A}$-optimality**: This criterion seeks to minimise the trace, the sum of the diagonal elements, of the inverse of the information matrix, denoted by $\text{tr}[\mathbf{M}^{-1}(\xi)]$. This means minimising the sum of the variance of the parameter estimators.

- **$\mathcal{V}$-optimality**: This criterion has been known by several other names; such as $\mathcal{I}$- and $\mathcal{Q}$- optimality criterion and is called the integrated or average variance

criterion. The $\mathcal{V}$-optimal design minimises the average prediction variance, $d(\xi)$, over the design region $\mathcal{X}$.

$$d(\xi) = \int_{\mathcal{X}} d(\boldsymbol{x}, \xi) d\boldsymbol{x}, \tag{3.4.5}$$

where $d(\boldsymbol{x}, \xi)$ is defined as in equation 3.4.4.

Despite the fact that there are many criteria in the design literature, until now the efficiency of a design has been mostly expressed in terms of $\mathcal{D}$-optimality (Kessels et al., 2009). This is because $\mathcal{D}$-optimum designs are easier to construct and consume less computer resources and time compared with the other criteria, particularly for Bayesian designs that require larger computer resources in any case. This follows since the criterion value for the $\mathcal{D}$-optimality is computed at the design points only and not averaged over the design region $\mathcal{X}$ as in, for example, **G**-optimality criterion where the time grows exponentially with the design region. In addition, the $\mathcal{D}$-optimality criterion has been observed to perform well in terms of other criteria, such as **G**-optimality and $\mathcal{V}$-optimality that are optimal for prediction (Kessels et al., 2004). However, one has to be aware that optimality of a design is related to the optimality criterion used to construct the design. Thus, a design that is optimal with respect to one criterion might not be optimal for another; for example, a $\mathcal{D}$-optimum design might not be $\mathcal{V}$-optimal.

Optimum designs are based on optimising the selected design criterion within the limitation of the search strategies to satisfy the design constraints. For most nonlinear models, specifically DCMs, the optimisation problem cannot be solved analytically, and hence a numerical optimisation procedure is required. This optimisation method uses different search algorithms which might result in different optimum design based on the search strategies used (see Section 3.7 for different choice design algorithms). Therefore, the optimisation procedure may result in a local, instead of global, optimum design, particularly when the optimisation function is not concave. The resulting design might be near optimum design but not necessarily the global optimum design. However, the term optimum is often used to represent near optimum design or to reflect a design

which is satisfactory from practical point of view.

In the optimal design literature, the optimality of different designs can be investigated using the general equivalence theorem introduced in the following section.

3.4.2 The General Equivalence Theorem

Kiefer and Wolfowitz (1960) prove that there exist ξ^* where the three following statements are equivalent:

- The design ξ^* maximises $|\mathbf{M}(\xi)|$.
- The design ξ^* minimises $\max_{x \in \mathcal{X}} d(\mathbf{x}, \xi)$.
- $\max_{x \in \mathcal{X}} d(\mathbf{x}, \xi^*) = k$.

This is known as the general equivalence theorem (GET) and it indicates that $\mathcal{D}$-optimum design is **G**-optimal. However, the key value of the theorem is that it states that the optimal design obtained by these criteria can be identified as having the maximum variance of the prediction equal to the number of the parameters in the model, k.

The optimality of different designs can be compared using the efficiency property introduced by Atkinson et al. (2007).

- **$\mathcal{D}$-efficiency**: This is used to compare the information content of an arbitrary design ξ to the optimum design ξ^*, and is defined as

$$\mathcal{D}_{eff} = \left\{ \frac{|\mathbf{M}^{-1}(\xi^*)|}{|\mathbf{M}^{-1}(\xi)|} \right\}^{1/k}. \tag{3.4.6}$$

  This measure shows how much better the optimal design is ξ^* compared to its competitor ξ.

- **Relative $\mathcal{D}$-efficiency**: This is used to compare the information content of any two designs. Thus, the relative $\mathcal{D}$-efficiency of design ξ_2 to ξ_1 is defined as

$$\mathcal{D}_{rel-eff} = \left\{ \frac{|\mathbf{M}^{-1}(\xi_1)|}{|\mathbf{M}^{-1}(\xi_2)|} \right\}^{1/k}. \tag{3.4.7}$$

  In this case, unlike the $\mathcal{D}$-efficiency, the relative efficiency value could be greater than one if design ξ_2 outperforms ξ_1 with respect to the optimality criterion value.

3.4.3 Example: Deriving the $\mathcal{D}$-optimal Design for the General Linear Model

In this section, we consider deriving the $\mathcal{D}$-optimum design for a simple linear model described in Atkinson et al. 2007, pp.44-57. The general form of the linear model can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}^T + \boldsymbol{\epsilon},$$

where $\mathbf{y}$ and $\boldsymbol{\epsilon}$ are a column vectors of the observations and the random errors, respectively, $\mathbf{X}$ is an $n \times k$ design matrix, and $\boldsymbol{\beta}$ is a vector of the unknown model parameters. We usually assume that the random errors are independent and identically normally distributed with mean zero and finite variance σ^2 (i.e. $E(\boldsymbol{\epsilon}) = 0$ and $Var(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}_n$, where $\mathbf{I}_n$ is $n \times n$ identity matrix).

There are different ways to fit the regression line; however, in this example we consider the most common method for fitting general linear model – ordinary least squares (OLS). In this method, the best fitting line for the observed data is computed through minimising the sum of squares of the error in fitting the lines, i.e. minimising the differences between each data point and the corresponding fitted value on the line. Assuming that $n > k$, then the minimisation process searches for the best estimators of $\boldsymbol{\beta}$ by solving k equations. If, as is usual, the random errors are assumed to be independent and identically normally distributed, the least squares estimators are the

same as the maximum likelihood estimators of the unknown parameters. Thus, in general, for the general linear model defined above, the estimators of the parameters can be written as

$$\hat{\boldsymbol{\beta}}^T = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y},$$

given that $(\mathbf{X}^T\mathbf{X})^{-1}$ exists. The variance-covariance matrix of the least squares estimator is given by

$$Var(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}.$$

If $\mathcal{D}$-optimality is to be considered as the design criterion, the design points must be selected such that the determinant of the variance-covariance matrix of the estimators is minimised (i.e. the confidence ellipsoid of the estimators is minimised), where the $\mathcal{D}$-optimality criterion is defined as

$$\mathcal{D} = |\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}|^{1/k}. \tag{3.4.8}$$

This is equivalent to maximising the determinant of the information matrix $\mathbf{X}^T\mathbf{X}$, where the higher the value of the determinant the lower the variance of the parameters will be that lead to higher design efficiency. To illustrate this consider computing $\mathcal{D}$-optimal design for a first-order linear model with one explanatory variable x and intercept as defined in equation 3.4.9:

$$y_i = \beta_0 + x_i\beta_1 + \epsilon_i \quad (i = 1, \ldots, n), \tag{3.4.9}$$

In matrix notation this can be written as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}^T + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is a $n \times 2$ design matrix, $\boldsymbol{\beta}^T = (\beta_0, \beta_1)^T$ and ϵ are a column vectors of unknown model parameters and random errors, respectively. If $n > 2$, and the random errors are assumed to be independent and identically normally distributed, the parameter estimates and the variance covariance matrix of the least squares estimators can be computed as defined

above, where

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}.$$

Thus, the determinant of the information matrix is

$$\begin{aligned} |\mathbf{X}^T\mathbf{X}| &= \begin{vmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{vmatrix} \\ &= nS_{xx} \end{aligned}$$

where $S_{xx} = \sum(x_i - \bar{x})^2$. For a given sample size $\mathcal{D}$-optimum design selects design points such that the determinant of the information matrix is maximised. Consequently, design points should be selected where S_{xx} is as large as possible. Therefore, for range of x over a finite interval $[x_{min}, x_{max}]$, the design points of a $\mathcal{D}$-optimum design must be chosen at the boundaries of that interval so the determinant of $\mathbf{X}^T\mathbf{X}$ is minimised.

From a frequentist perspective, this $\mathcal{D}$-optimum design minimises the volume of the confidence ellipsoid of all parameters. This is because the volume of the confidence ellipsoid of the parameters is inversely proportional to the determinant of the information matrix $\mathbf{X}^T\mathbf{X}$. This ellipsoid is derived from the $100(1-\alpha)\%$ confidence region for all parameters that has the following form

$$(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\mathbf{X}^T\mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \leq ks^2F_{k,v,\alpha},$$

where k is the total number of parameters, s^2 is the estimator of σ^2 on v degrees of freedom and $F_{k,v,\alpha}$ is the $\alpha\%$ point of the F distribution on k and v degrees of freedom. From this equation, the volume of the confidence ellipsoid can be derived as

$$\frac{(\pi ks^2F_{k,v,\alpha})^{k/2}}{\Gamma(k/2+1)|\mathbf{X}^T\mathbf{X}|^{1/2}},$$

where$\Gamma(k/2+1) = (k/2)!$. Therefore, the $\mathcal{D}$-optimum design minimises the volume of the confidence ellipsoid, through maximising the determinant of the information matrix, $|\mathbf{X}^T\mathbf{X}|$.

In a Bayesian approach, $\mathcal{D}$-optimality can be interpreted as minimising the volume of the credible ellipsoid of the unknown model's parameters, i.e. minimising the determinant of the posterior variance-covariance matrix (VCM). For a normal linear model with conjugate priors for the mean and the variance, this VCM would have a closed form and would not be depend on the unknown model parameters. Therefore the Bayesian optimum design would be simply obtained by minimising this VCM over the design points, as in the non-Bayesian design, to return the required design with smallest values for the posterior VCM (see Atkinson et al., 2007; and Chaloner and Verdinelli, 1995, for more details about Bayesian design for linear models).

However, for nonlinear models, unlike linear models, the VCM of the estimators, and hence the design criterion, depends on the unknown model parameters (Atkinson et al., 2007). This complicates the computation of the optimality criterion, and consequently the construction of the optimum design for nonlinear model, here DCMs. Recently, however, the Bayesian approach has become the most usual method for coping with the problem of design dependency on the unknown parameters of the choice model. This approach and the construction of Bayesian optimal design for DCMs are discussed in the following section.

3.5 Nonlinear Optimal Design Problem

Calculating optimal designs for nonlinear models, particularly DCMs, is more complicated than for linear models. This is because the design criterion is usually a function of the information matrix of the design or its inverse which for nonlinear model itself depends on the specification and parameter values of the model, denoted as $\mathbf{M}(\xi, \boldsymbol{\beta})$. For instance, Kessels et al. (2004) showed that the information matrix $\mathbf{M}(\xi, \boldsymbol{\beta})$ for the multinomial logit model (MNL) model depends on the unknown model parameters through the choice probability as

$$\mathbf{M}(\xi, \boldsymbol{\beta}) = \sum_{s=1}^{S} \mathbf{X}_s(\mathbf{P}_s - \mathbf{p}_s\mathbf{p}_s^T)\mathbf{X}_s^T, \tag{3.5.1}$$

where $\mathbf{X}_s$ is the $k \times J$ design matrix for each choice set s, k is the number of parameters, and the concatenated $\mathbf{X}_s$ matrices constitute the design matrix $\mathbf{X}$ of the choice experiment. $\mathbf{P}_s$ and $\mathbf{p}_s$ are a diagonal matrix and a $J \times 1$ vector of the choice probabilities of each alternative presented in the choice set s which is a function of the unknown model parameters as $P_{js} = \frac{e^{\boldsymbol{\beta}\boldsymbol{x}_{js}^T}}{\sum_{t=1}^{J} e^{\boldsymbol{\beta}\boldsymbol{x}_{ts}^T}}$.

A Bayesian approach has been used to circumvent such dependency problems through incorporating prior information regarding the parameter values into the experimental design. In the design literature, there are two approaches to incorporating the prior information about the unknown model parameters, which are either: using prior point estimates (which results in what are called locally optimal designs), or using the entire prior distribution (that results in what are called Bayesian optimal designs) as discussed in the following sections.

In this thesis, we consider using a Bayesian optimal design approach to deal with the dependency of the optimal choice design on the values of the unknown model parameters while assuming that the specified model is correct (i.e. the dependency of the optimal design on the specification of the model is ignored here by assuming the underlying model is the correct model).

3.5.1 Locally Optimal Designs

As mentioned earlier, for nonlinear models, the information matrix, or any function of it, depends on the model parameters, and consequently so does the optimality criterion and the construction of the choice design. The local optimal design is constructed by optimising the criterion value given the best prior guess for the parameter values, $\boldsymbol{\beta}$. This approach was first introduced to DCEs in Bunch et al. (1996) who assume zero prior values for the preference parameters $\boldsymbol{\beta}_0$, which results in what are called utility-neutral designs. However, this assumption is unrealistic, as it implies respondents have no preference over all attribute levels, and hence all alternatives, for any possible choice questions. Huber and Zwerina (1996) argue that this assumption might be costly and result in inefficient choice designs, particularly when the true parameter values are not zero, as is the case in most practical marketing choice studies. In their study, they introduced non-zero prior point estimates, $\boldsymbol{\beta}_p$, to calculate the local optimal choice designs. It can be noticed that the utility-neutral design is a special case of the local optimal designs with zero prior point estimate.

In local optimal design, researchers usually use the optimality criterion with subscript p to refer to the point prior estimate approach, to distinguish it from the criterion that is used for linear model, where no prior is required. For example, for $\mathcal{D}$-optimality a criterion is denoted as $\mathcal{D}_p$ when a non-zero prior point estimate is used, and for the special case of a zero prior it is denoted by $\mathcal{D}_0$. The locally $\mathcal{D}$-optimum design is calculated by maximising the information matrix of the DCM model $\mathbf{M}(\xi, \beta)$, or equivalently minimising its inverse, given the prior point estimates. Formally, the design

criterion is defined as

$$\mathcal{D}_p = |\mathbf{M}^{-1}(\xi, \boldsymbol{\beta_p})|^{1/k}, \tag{3.5.2}$$

and, for the zero prior case, as

$$\mathcal{D}_0 = |\mathbf{M}^{-1}(\xi, \boldsymbol{\beta_0})|^{1/k}. \tag{3.5.3}$$

The prior information about the parameters could be obtained from previous studies, pilot studies, or expert judgments. A zero prior is typically used to simplify the choice design problem. However, as mentioned earlier, this assumption is unrealistic and impractical, and might reduce the efficiency of the designs.

A more general problem with this approach is that any point estimate does not take into account the uncertainty around the assumed parameter value, and a poor guess for the prior point estimate may result in inefficient choice design. Therefore, in the next section, we introduce a more advanced experimental design approach, Bayesian optimum design, that allows researchers to account for the uncertainty by assuming a prior distribution for the possible parameter values.

3.5.2 Bayesian Optimal Designs

As has been discussed, the efficiency of locally optimal designs depends on the choice of the point prior: a poorly defined prior leads to inefficient choice designs. Therefore, more recently, Bayesian optimum designs have been widely used in the design literature to provide a more robust design solution. In this approach, the dependence on single values is avoided by using a prior distribution for the unknown parameter vector, denoted by $\pi(\boldsymbol{\beta})$, and, hence, constructing the optimum design based on several plausible parameter values. This approach might lead to a more informative design, as it accounts for uncertainty in the parameter values. One should note that locally optimal design is a special case of Bayesian design, with point priors either zero, $\boldsymbol{\beta}_0$,

or non-zero, $\boldsymbol{\beta}_p$.

This approach was first introduced to DCEs in Sándor and Wedel (2001) who showed an improvement in constructing optimal choice designs for the MNL model using a manager's prior belief over the local optimal design, particularly, with poor guesses for the point estimates. This approach was then implemented by many researchers in different fields to construct Bayesian optimum designs for DCEs (e.g. Kessels et al., 2004, Kessels et al., 2006, Rose and Bliemer, 2008, Kessels et al., 2011).

The approach requires a prior distribution for the unknown model parameters. In general, the multivariate normal distribution, $\mathcal{N}(\boldsymbol{\beta}|\boldsymbol{\mu}, \Sigma)$, has been used, though it might not be appropriate in some situations (e.g. when the price is included as an attribute in the experiments, or for parameters associated with decrement in ordered attributes). As a result, more attention should be given to selecting the prior distribution of the parameters so that it reflects, for example, the correct sign of the parameter values, and the expected size of effects. The prior information about the sign or the values of the unknown parameters could be collected from previous studies, pilot studies, expert judgments and other method such as sequential design strategies discussed in Carlsson and Martinsson (2003).

A Bayesian optimal design is calculated by optimising an appropriate function of the information matrix or its inverse over a prior distribution of the parameter values. For example, a Bayesian $\mathcal{D}$-optimum design can be computed by minimising the expectation of the determinant of the inverse of the information matrix, $\mathbf{M}(\xi, \boldsymbol{\beta})$, that is given by

$$\begin{aligned} \mathcal{D}_B &= E_{\boldsymbol{\beta}}|\mathbf{M}^{-1}(\xi, \boldsymbol{\beta})|^{1/k} \\ &= \int |\mathbf{M}^{-1}(\xi, \boldsymbol{\beta})|^{1/k} \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}, \end{aligned} \tag{3.5.4}$$

where the subscript B is used to refer to Bayesian optimality criterion, reflecting the use of a Bayesian approach.

From a Bayesian perspective, both $\mathcal{D}_p$-optimality and $\mathcal{D}_B$-optimality criteria should minimise the expected posterior VCM of the unknown parameters. Therefore, the corresponding design criterion should be formulated using the exact posterior VCM of the unknown parameters in order to generate Bayesian optimal designs in a true Bayesian approach. However, for nonlinear models in general, and DCMs particularly, this matrix does not have closed form, and, hence, neither does the optimality criterion. Thus, Bayesian optimal designs are typically based on different approximations of the posterior VCM instead, as described in the following section.

3.6 Bayesian Experimental Design Criteria for Nonlinear Models

As stated above, constructing Bayesian optimum designs for nonlinear models requires calculating the posterior VCM of the unknown model parameters to form the Bayesian design criterion. However, this matrix cannot be derived analytically for the DCMs, since it depends on the unknown parameters of the choice model and an alternative approach must be sought. In the design literature, there are different approximation methods, which result in different forms for the design criterion.

Berger (1985) gives several asymptotic approximations for the posterior VCM; however, in this section, we mainly focus on two approximation methods described in Yu et al. (2008). The first method approximates the expected posterior VCM based on asymptotic theory using either the second derivative of the log likelihood function, the Fisher information matrix, or the second derivative of the log posterior density of

model parameters, the generalised Fisher information matrix. This results in what is called asymptotic Bayesian design criteria as illustrated in Section 3.6.1. The alternative approach does not depend on asymptotic theory. It is based on a true Bayesian approach using the exact expected posterior covariance matrix that leads to an exact Bayesian criterion, as in Section 3.6.2, or a Bayesian information approach, as described in Section 3.6.3.

3.6.1 Asymptotic Bayesian Criteria

The posterior distribution of the choice model parameters is often approximated using an asymptotic normal approximation of the maximum likelihood estimators (m.l.e.). This approximation assumes that the posterior distribution looks increasingly like a normal distribution as the sample size become larger (Train, 2003, pp.284–291). The VCM of this asymptotic distribution is approximated by the inverse of the Fisher information matrix (FIM) of design ξ as

$$\widehat{Var}(\boldsymbol{\beta}) = \text{FIM}^{-1}(\xi, \boldsymbol{\beta}), \tag{3.6.1}$$

where the FIM is defined as the negative (value) of the expectation of the Hessian matrix, $\mathbf{H}$, or the second derivative of the log likelihood function of the underlying model, denoted, in general, as $L(\mathbf{y}|\boldsymbol{\beta})$:

$$\begin{aligned} \text{FIM}(\xi, \boldsymbol{\beta}) &= -E_{\mathbf{Y}}\left[\mathbf{H}(\xi, \boldsymbol{\beta})\right], \\ &= -E_{\mathbf{Y}}\left[\frac{\partial^2 \log L(\mathbf{y}|\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2}\right]. \end{aligned} \tag{3.6.2}$$

Thus, the information in a design ξ is quantified as the negative value of the expected derivative of the score function $\frac{\partial \log L(\mathbf{y}|\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ as defined in Equation (3.6.2). An alternative for the FIM approximation, also rooted in asymptotic theory, is the generalised Fisher information matrix (GFIM), in which the expected posterior variance is approximated

as the inverse of the GFIM of the design, that is computed as the negative value of the expectation of the second derivative of the log posterior density, $\pi(\boldsymbol{\beta}|\mathbf{y}) \propto L(\mathbf{y}|\boldsymbol{\beta})\pi_1(\boldsymbol{\beta})$, as

$$\begin{aligned} \text{GFIM}(\xi, \boldsymbol{\beta}) &= -E_{\mathbf{Y}} \left\{ \frac{\partial^2 \log \left[L(\mathbf{y}|\boldsymbol{\beta})\pi_1(\boldsymbol{\beta}) \right]}{\partial \boldsymbol{\beta}^2} \right\}, \\ &= -E_{\mathbf{Y}} \left\{ \frac{\partial^2 l(\mathbf{y}|\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} \right\} - E_{\mathbf{Y}} \left\{ \frac{\partial^2 \log \pi_1(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} \right\}, \\ &= \text{FIM}(\xi, \boldsymbol{\beta}) - E_{\mathbf{Y}} \left\{ \frac{\partial^2 \log \pi_1(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} \right\}, \end{aligned} \tag{3.6.3}$$

where the FIM is defined as in equation (3.6.2), $\pi_1(\boldsymbol{\beta})$ is the inference prior, and $-E_{\mathbf{Y}} \left\{ \frac{\partial^2 \log \pi_1(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} \right\}$ is the amount of information that the prior carries about the parameters. Based on asymptotic theory, this is equal to the inverse of the covariance matrix of the prior distribution; thus, supposing that the prior distribution follows a multivariate normal distribution with covariance matrix Σ,

$$E_{\mathbf{Y}} \left\{ \frac{\partial^2 \log \pi_1(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} \right\} = -\Sigma^{-1}.$$

The GFIM therefore becomes

$$\text{GFIM}(\xi, \boldsymbol{\beta}) = \text{FIM}(\boldsymbol{\beta}, \xi) + \Sigma^{-1}. \tag{3.6.4}$$

The asymptotic Bayesian design criteria are formulated using the asymptotic approximation of the posterior VCM using the FIM and GFIM. However, both the FIM and the GFIM depend on the unknown parameters of the choice model. As a result, optimum designs are constructed in this framework using an appropriate prior distribution, design prior denoted as $\pi_2(\boldsymbol{\beta})$, to take into account the uncertainty in the possible parameter values. Using the asymptotic approximation of the expected posterior VCM, the $\mathcal{D}$-optimality criterion is computed as

$$\mathcal{D}^B_{\text{FIM}} = \int |\text{FIM}^{-1}(\xi, \boldsymbol{\beta})|^{1/k} \pi_2(\boldsymbol{\beta}) d\boldsymbol{\beta}, \tag{3.6.5}$$

or

$$\mathcal{D}^{B}_{\text{GFIM}} = \int |(\text{FIM}(\xi, \boldsymbol{\beta}) + \Sigma^{-1})^{-1}|^{1/k} \pi_2(\boldsymbol{\beta}) d\boldsymbol{\beta}. \tag{3.6.6}$$

Note that in Equation (3.6.6), different priors could be used for the inference (estimating the posterior VCM), $\pi_1(\boldsymbol{\beta})$, and the design prior (constructing Bayesian design), $\pi_2(\boldsymbol{\beta})$. However, using one prior, $\pi(\boldsymbol{\beta})$, for both is more sensible, as analysis and design are part of the same overall study. Also, the FIM criterion is just a special case of the GFIM, as the later converges to the FIM when assuming a non-informative prior distribution in which Σ^{-1} becomes close to zero (e.g. if the prior is a normal distribution with large variance).

These asymptotic approximations of the posterior VCM rely on large-sample theory, i.e. a large sample size is required to ensure the validity of the asymptotic approximation. McCulloch and Rossi (1994) state that a large sample size, perhaps more than 1,000 observations per model parameter, is essential to ensure the validity of the asymptotic approximations of the posterior distribution. Nevertheless many researchers in the choice design literature use Bayesian asymptotic criteria approximated using the FIM as a basis for constructing efficient choice design due to its computational simplicity (e.g. Huber and Zwerina 1996; Zwerina et al. 1996; Sándor and Wedel 2001; and Kessels et al. 2006). However, in a situation where sample size is an issue, the FIM might result in a poor approximation of the posterior VCM, and consequently result in inefficient designs. In this case, it would be better to use another approximation method such as the GFIM that is known to have better finite sample properties than the FIM, or even better use a better approximation for the true posterior distribution, as discussed in the following section.

3.6.2 Exact Bayesian Criteria

This method uses the exact value of the expected posterior variance-covariance matrix (EPVC), and does not depend on the asymptotic theory. In this approach, the EPVC

is estimated by the expected posterior covariance matrix of the Bayesian estimators after conducting the experiment. Here, and assuming that the prior for the inference is the same as the one for the design (i.e. $\pi_1 = \pi_2$), the posterior density of $\boldsymbol{\beta}$ given the data $\mathbf{y}$ is given by

$$\pi(\boldsymbol{\beta}|\mathbf{y}) = \frac{L(\mathbf{y}|\boldsymbol{\beta})\pi(\boldsymbol{\beta})}{p_{\mathbf{Y}}(\mathbf{y})}, \tag{3.6.7}$$

where $p_{\mathbf{Y}}(\mathbf{y})$ is the marginal distribution of the observations $\mathbf{y}$, and $\pi(\boldsymbol{\beta})$ is the prior distribution of the parameters. Thus, the posterior VCM of the parameter estimators is given by

$$\begin{aligned} Var(\boldsymbol{\beta}|\mathbf{y}) &= \int (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})^T \pi(\boldsymbol{\beta}|\mathbf{y}) d\boldsymbol{\beta}, \\ &= \frac{1}{p_{\mathbf{Y}}(\mathbf{y})} \int (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})^T L(\mathbf{y}|\boldsymbol{\beta})\pi(\boldsymbol{\beta}) d\boldsymbol{\beta}, \end{aligned} \tag{3.6.8}$$

where $\bar{\boldsymbol{\beta}}$ is the posterior mean, given by

$$\bar{\boldsymbol{\beta}} = \int \boldsymbol{\beta}\, \pi(\boldsymbol{\beta}|\mathbf{y}) d\boldsymbol{\beta}.$$

Note that the posterior VCM depends on the response values $\mathbf{y}$ (here choices), which will not be observed before conducting the experiment. Therefore, the posterior variance of the estimators is calculated by taking the expectation of the posterior variance in Equation (3.6.8) over the marginal distribution of the responses, $p_{\mathbf{Y}}(\mathbf{y})$, as follows:

$$\begin{aligned} E\left[Var(\boldsymbol{\beta}|\mathbf{y})\right] &= \int Var(\boldsymbol{\beta}|\mathbf{y})\, p_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y}, \\ &= \int \left[\frac{1}{p_{\mathbf{Y}}(\mathbf{y})} \int (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})^T L(\mathbf{y}|\boldsymbol{\beta})\pi(\boldsymbol{\beta}) d\boldsymbol{\beta} \right] p_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y}, \\ &= \int \int (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})^T L(\mathbf{y}|\boldsymbol{\beta})\pi(\boldsymbol{\beta})\, d\mathbf{y}\, d\boldsymbol{\beta}. \end{aligned}$$

Hence finding an exact Bayesian $\mathcal{D}$-optimum design requires minimising the determi-

nant of the EPVC of the parameter estimators, that is

$$\mathcal{D}_{EPVC} = \left| \int \int (\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})^T L(\mathbf{y}|\boldsymbol{\beta}) \pi(\boldsymbol{\beta}) d\mathbf{y} d\boldsymbol{\beta} \right|^{1/k}. \tag{3.6.9}$$

A design that minimises Equation (3.6.9) is called an exact Bayesian $\mathcal{D}$-optimum design.

3.6.3 Bayesian Information Criteria

The Bayesian information approach, which computes the gain in knowledge about Bayesian estimators when moving from the prior distribution to the posterior distribution, was introduced to the experimental design field by Lindley (1956). The information design criterion is based on the concept of the Shannon information introduced by Shannon (1948) to measure the uncertainty associated with a random variable. Using this concept, an optimal design must maximise the expected gain in Shannon information or, equivalently, the amount of information provided by a design.

In a Bayesian framework, the gain in Shannon information is calculated as the difference between the information provided by the posterior distribution, denoted by $g_1(\mathbf{y})$, and that provided by the prior distribution, denoted by g_0. Formally, the amount of information provided by the prior distribution is given by

$$\begin{aligned} g_0 &= E_{\boldsymbol{\beta}} \{\log[\pi(\boldsymbol{\beta})]\} \\ &= \int \log[\pi(\boldsymbol{\beta})] \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}. \end{aligned} \tag{3.6.10}$$

The posterior Shannon information, obtained after conducting the experiment and

observing the response vector $\mathbf{y}$, is given by

$$\begin{aligned} g_1(\mathbf{y}) &= E_{\boldsymbol{\beta}}\left\{\log[\pi(\boldsymbol{\beta}|\mathbf{y})]\right\} \\ &= \int \log[\pi(\boldsymbol{\beta}|\mathbf{y})]\pi(\boldsymbol{\beta}|\mathbf{y})d\boldsymbol{\beta} \\ &= \frac{1}{p_{\mathbf{Y}}(\mathbf{y})}\int \log[\pi(\boldsymbol{\beta}|\mathbf{y})]L(\mathbf{y}|\boldsymbol{\beta})\pi(\boldsymbol{\beta})d\boldsymbol{\beta}. \end{aligned} \tag{3.6.11}$$

Note that, again, the posterior Shannon information depends on the response vector $\mathbf{y}$ which has not been observed yet at the design stage. Therefore, the expected Shannon information is computed instead. Thus,

$$\begin{aligned} g(\pi(\boldsymbol{\beta}), \mathbf{y}) &= E_{\mathbf{Y}}[g_1(\mathbf{y}) - g_0] \\ &= \int [g_1(\mathbf{y}) - g_0]\, p_{\mathbf{Y}}(\mathbf{y})d\mathbf{y}. \end{aligned} \tag{3.6.12}$$

The optimal design should maximise the expected Shannon information gain provided in Equation (3.6.12). However, the amount of information provided by the prior distribution, g_0, does not depend on the the design. Therefore, it can be ignored, and, hence, the Bayesian optimality criterion based on Shannon information can be defined as

$$\mathcal{D}_{Shannon} = \int\int \log\left[\pi(\boldsymbol{\beta}|\mathbf{y})\right] L(\mathbf{y}|\boldsymbol{\beta})\pi(\boldsymbol{\beta})d\mathbf{y}d\boldsymbol{\beta}. \tag{3.6.13}$$

The EPVC and Shannon information criteria, unlike the asymptotic criteria, do not rely on asymptotic theory. Thus, their approximation to the exact VCM of the model parameters and the amount of information gained, respectively, is often valid no matter what sample size is available (e.g. number of respondents or choice questions). Also, both criteria take into account the prior knowledge about the model parameters in the design and estimation stages, thus providing efficient designs for a fully Bayesian framework, whereas the asymptotic Bayesian criteria consider the prior information for the design phase only (Yu et al., 2008).

Nevertheless, the EPVC and Shannon information criteria value depend on the

response vector, $\mathbf{y}$, that has not been observed yet at the design phase. This means that using these criteria as a basis for constructing optimal designs comes at large computational cost, as they require an intensive computational integration over the marginal distribution of the responses. This makes them less practical.

Therefore, for computational simplicity, the asymptotic criteria will be used as the basis for constructing Bayesian choice designs throughout this thesis. Constructing Bayesian designs based on even these simpler criteria requires a numerical/algorithmic approach to optimise the criterion over a suitable prior distributions of the parameters, and often involves complicated integrals. In the following section, we review the algorithms and software available to construct Bayesian choice designs.

3.7 Design Algorithms and Software

The computation of the optimality criteria and hence selection of the optimal choice designs requires search algorithms and computer software to identify the design which best satisfies the study constraints. In this section, we review the most widely used algorithms and design software available to calculate locally optimal designs and Bayesian optimal choice designs for discrete choice models in Sections 3.7.1 and 3.7.2, respectively. An investigation of the ability of these algorithms and design software, particularly ones for Bayesian designs, to accommodate our design model and constraints is provided in Chapter 5; here we mainly focus on recent developments in Bayesian optimal choice design algorithms.

3.7.1 Local and Utility-neutral Optimal Design Algorithms

In the previous sections, we illustrated that the calculation of the optimal design criterion and hence the choice design depends on the unknown model parameters; therefore prior knowledge is required to construct the choice design. In the Bayesian choice design literature this is mainly done by constructing locally optimal designs using the best prior point estimates. Either a noninformative point prior (i.e. assuming all the preference parameters are zero) or non-zero point prior is used to construct the optimal choice design. In this section, we review design algorithms used to generate utility-neutral and local optimal designs.

Anderson and Wiley (1992) argue that design strategies used to construct optimal designs based on $\mathcal{D}$-optimality criteria for linear models can work well for nonlinear choice models ignoring the dependence of the optimality criterion on the unknown model parameters by assuming zero priors. This is because this assumption simplifies the optimisation problem and reduces it to a linear design problem. They provide a catalogue of utility-neutral designs for the MNL choice models based on the $\mathcal{D}$-optimality criterion for linear models. The catalogued designs enable users to estimate the attribute effects.

Nevertheless, this catalogue might not be efficient to construct a design for a real choice design problems, since, in practice, most choice design problems are complicated and might required larger designs than those covered by the catalogue. Therefore, Kuhfeld et al. (1994) recommended the use of computerised search algorithms to find optimal designs for discrete choice models. They reviewed general computerised approaches for generalised linear models built in SAS software.

In their study, they reviewed the Dykstra (1971) algorithm and the DETMAX algorithm of Mitchell (1974) which are faster than, but not as efficient as, the Mitchell and Miller (1970) simple exchange algorithm. The DETMAX and simple exchange algorithms require an initial design, while the Dykstra (1971) algorithm starts with an

empty design and sequentially adds design points from a pre-defined candidate set (a set of all possible attribute level combinations that can be included in the design) in which the design criteria is optimised. These algorithms then all improve the efficiency of the current proposed design by adding a candidate point and deleting a design point one at a time. Kuhfeld et al. (1994) make use of the Cook and Nachtsheim (1980) algorithm, which is a modified version of the Federov (1972) exchange algorithm. Both these algorithms require a random starting choice design, and are based on an exchange procedure that adds a candidate point and deletes a design point simultaneously. This makes them more reliable, but slower, in finding the optimal design than the earlier algorithms. The modified algorithm speeds up the exchange procedure and hence the search for $\mathcal{D}$-optimum design by executing any beneficial exchange as soon as it is discovered rather than only performing the best exchange. Therefore, the modified design is as efficient as the simple Federov (1972) algorithm in terms of finding the optimal design, but is much faster in finding the best candidate point from the candidate set to switch with each point in the starting design (Kuhfeld et al., 1994).

All these search algorithms are based on classical linear design principles that ignore the nonlinear nature of the choice models. This reduces the efficiency of their choice designs. Therefore, Huber and Zwerina (1996) identified four principles that they believed to be characteristic of an optimal choice design, i.e. satisfying these principles jointly produces optimal choice designs. These principles are level balance, orthogonality, minimal overlap and utility balance, described in Section 3.2.3. Their algorithm uses orthogonal arrays to generate the first alternative in each choice set, via the OPTEX procedure in SAS, and then uses the shift/foldover procedure, first developed by Louviere et al. (2000), to produce the subsequent alternatives in each choice set. The resultant design has perfect level balance, orthogonality, and minimal overlap, as the foldover procedure generates alternatives that imitate the perfect level balance and orthogonality of the initial array. Designs satisfying these three properties provides an efficient utility-neutral design or $\mathcal{D}_0$-optimum design, since the fourth principle, utility balance, is intuitively satisfied by assuming $\boldsymbol{\beta} = \mathbf{0}_k$.

Huber and Zwerina (1996) argue that the assumption of zero prior parameter values might be inappropriate for many real design problems, especially when there is some knowledge available about the model coefficients. They introduce two methods to improve the utility balance of the orthogonal design or utility-neutral design when the parameter values are not zero. These are swapping and relabelling methods, as described in 3.7.2. This algorithm is known as the RS-algorithm and is also written in SAS language.

The study illustrated that relabelling improves the utility balance property without affecting other design properties, whereas swapping can result in a better utility balanced design by sacrificing some orthogonality and degrading the $\mathcal{D}_0$ criterion value. They show that, for a design of 15 pairs with 4 attributes each with 3 levels, denoted as $3^4/2/15$, the swapping procedure raises the design efficiency by 10% (this improvement might exceed 50% for more complex designs) and degrades the $\mathcal{D}_0$ criterion value by 32%. Therefore, they recommend using the relabelling over the swapping procedure and selecting the best relabelling design instead of the swapping design, or an algorithm that includes both procedures, if the researchers are not sure whether $\boldsymbol{\beta} \neq 0$.

However, as discussed in Section 3.2.3, in most design problems these principles cannot be jointly satisfied, and even if they were it does not guarantee that the design is optimal (Street and Burgess, 2007). In addition, generating choice designs directly from these principles still follows linear design principles, and hence might not optimise the correct criterion for the choice model. Therefore, Zwerina et al. (1996) recommended using computerised search algorithms that construct efficient choice designs by directly optimising the correct design criterion for the MNL, while allowing the user to incorporate any anticipated parameter values instead of assuming zero priors or generating optimal design directly from the formal principles. In their work, they extend the work of Kuhfeld et al. (1994) and modified their search algorithm, itself a modification of the Federov (1972) exchange algorithm, to account for the correct design criterion corresponding to the MNL model and incorporate any anticipated parameter values in

the optimisation process. This new version of the exchange algorithm is known as the modified Fedorov choice algorithm which built in SAS/IML program.

The modified Fedorov choice algorithm, like the Kuhfeld et al. (1994) algorithm, requires a random starting choice design and pre-specified candidate set of all possible alternatives. The exchange procedure starts with replacing the first alternative in the choice design with the alternative from the candidate set that optimises the $\mathcal{D}$ value. The procedure is repeated with all alternatives in the design and continued until no more improvement is possible. The optimisation search procedure must be repeated for different starting designs, storing the best design among these tries, in order to avoid poor local optima. The resulting design, unlike the RS optimal design, is not restricted to the optimal design principles (level balance, orthogonality, minimal overlap and utility balance). Also, the modified Fedorov algorithm is more general and can be applied to any level of design complexity, which might not exist with the orthogonal arrays in the RS algorithm.

Several studies have since been conducted to develop the optimal theory of choice designs, particularly for the MNL model. Kessels et al. (2009) describe the work conducted by Street et al. (2001), and Street and Burgess (2003; 2004) who constructed $\mathcal{A}_0$-optimum choice designs for the MNL model and generated $\mathcal{D}_0$-optimum designs for experiments with two-level attributes for any equal choice set size. This was then extended by Street and Burgess (2005) to generate $\mathcal{D}_0$-optimum designs for the MNL model for attributes with any number of levels. In general, Street and Burgess's designs are constructed using a shifting procedure applied to a starting design based on an orthogonal array, where they shift the first profile in each choice question to create the subsequent alternatives, as in the Huber and Zwerina (1996) algorithm. Refer to Street and Burgess (2007) for more details about Street and Burgess's design algorithms.

As this thesis mainly considers Bayesian choice design using a prior distribution, in the following section we review the development in the design algorithms that allow incorporation of a prior distribution for the unknown model parameters.

3.7.2 Bayesian Optimal Design Algorithms

Local optimal designs do not take into account the uncertainty in the assumed parameter values; however, Bayesian designs deal with this problem by assuming a prior distribution for the unknown parameters and optimising the design over this prior belief. This approach was introduced in Sándor and Wedel (2001) who constructed a Bayesian $\mathcal{D}$-optimum design for the MNL model using the relabelling and swapping (RS) algorithm developed in Huber and Zwerina (1996) in addition to another procedure called cycling (C) to construct what is called the RSC algorithm. However, the swapping procedure in the RSC algorithm is slightly different from the one used by Huber and Zwerina (1996). The RSC algorithm is written in the GAUSS programming language. Here, we explain this algorithm briefly; refer to Sándor and Wedel (2001) for simple examples and more detail about each procedure in the algorithm.

- **Relabelling:** In the relabelling procedure the levels of the attributes are permuted across choice sets searching for a combination of permutations that gives the best design with the highest efficiency. The procedure starts with the first attribute in the first choice set and passes that through all attributes and choice sets. So, returning to AQL-5D example, the method involves an investigation of $5! \times 5! \times 5! \times 5! \times 5! = 120^5$ possible designs, as each of the attributes has five levels, and hence $5! = 120$ possible permutations. The method returns the best possible reliable design, i.e. the relabelled design with the smallest $\mathcal{D}_B$ value. This design is called the optimal R-design.

- **Swapping:** This involves switching two attribute levels between alternatives within a choice set. Further, this procedure, unlike the swapping procedure developed in Huber and Zwerina (1996), considers simultaneously swapping the levels of several attributes within a choice set.

  Thus the procedure starts with the first choice set of the best relabelled design, and swaps the attribute levels of the first attribute in the first alternative with

the level of this attribute in the second alternative. The change is executed if it improves the criterion value. The same procedure is passed to the following choice set where the swapping procedure will start over if an improvement occurred. For instance, consider the following choice sets from the AQL-5D classification system: the procedure starts with the first choice set and swaps the attribute levels of the first attribute between alternatives and then compares the value of the $\mathcal{D}$ optimality criterion.

$$\underset{\mathcal{D}_1}{\left[\begin{array}{ccccc} 1 & 2 & 3 & 2 & 4 \\ 2 & 1 & 0 & 2 & 4 \\ \hline 0 & 1 & 2 & 1 & 0 \\ 1 & 2 & 0 & 1 & 0 \end{array}\right]} \longrightarrow \underset{\mathcal{D}_2}{\left[\begin{array}{ccccc} 2 & 2 & 3 & 2 & 4 \\ 1 & 1 & 0 & 2 & 4 \\ \hline 0 & 1 & 2 & 1 & 0 \\ 1 & 2 & 0 & 1 & 0 \end{array}\right]}.$$

If $\mathcal{D}_2 < \mathcal{D}_1$ then the change is executed, and move to swap the attribute levels of this attribute in the second choice set as follows:

$$\underset{\mathcal{D}_3}{\left[\begin{array}{ccccc} 2 & 2 & 3 & 2 & 4 \\ 1 & 1 & 0 & 2 & 4 \\ \hline 1 & 1 & 2 & 1 & 0 \\ 0 & 2 & 0 & 1 & 0 \end{array}\right]}.$$

If $\mathcal{D}_3 < \mathcal{D}_2$ then the exchange is executed, and the swapping procedure should start over again from the first choice set and so on until no more improvement is possible. The swapping continues through all attributes an choice sets until all swaps and simultaneous swaps have been investigated. When all choice sets have been examined, the best modified design returned is called the optimal RS-design.

- **Cycling:** This algorithm involves two procedures: cyclically rotating the attribute levels (e.g. for an attribute with three levels, level 1 is replaced by level 2, level 2 by level 3, and level 3 by level 1), and swapping them. Thus it starts with cyclically rotating the levels of the first attribute in the first choice set of the

optimal RS-design until all possibilities are examined, and adopts the one that most improves the design criterion. Swapping is applied then to the attribute levels among alternatives within that choice set, and once again cycles through a rotation of the attribute levels. The algorithm moves to the first attribute in the second choice set and continues until the last choice set. The same procedure is repeated for the remaining attributes until all cycles and swaps are verified. At each stage of the cycling algorithm, if an improvement occurs then the change is returned and the procedure starts all over again from the first attribute in the first choice set. The procedure will stop if there is no further possible improvement. The last modified design is called the optimal RSC-design.

The Sándor and Wedel (2001) study illustrates that the $\mathcal{D}_B$-optimum design is more efficient than the locally $\mathcal{D}_p$-optimum design, especially when there is a large uncertainty in the assumed parameter values. However, when the prior point estimates of the parameter values are close to the real values or the value of the parameters are known for certain, then the locally RS optimal design generated by Huber and Zwerina (1996) tends to perform better. On the other hand, Bayesian designs are still more robust against a poor initial guess for the parameter vector than non-Bayesian ones (Yu et al., 2008).

The RSC algorithm requires a starting design that satisfies level balance and the minimal overlap properties. An updated version of the RSC algorithm has been developed by Sándor and Wedel (2002). In the new version, they modify the cycling procedure: they replace the combination of the cyclically rotating the attribute levels and swapping them by cycling the attribute level through the choice design. Also, the modified RSC algorithm is not restricted to the level balance and the minimal overlap properties. This makes the algorithm more amenable to design improvements.

Since the introduction of Bayesian design in 2001, the Bayesian approach has been increasingly used to cope with the problem of design dependence on unknown parameter values. Many studies have been carried out to develop different design algorithms

to produce optimal Bayesian choice designs. Kessels et al. (2004) start with adapting the modified Fedorov choice algorithm developed in Zwerina et al. (1996) in a Bayesian framework to account for the uncertainty on the prior point estimate through incorporating the prior distribution of the parameters in the optimisation procedure. The new version of the algorithm is available in SAS/ChoicEFF as a macro, and is known as the Bayesian modified Fedorov choice algorithm or Monte Carlo modified Fedorov (MCMF) algorithm, since the design criterion is approximated using Monte Carlo samples from the parameter distribution. This algorithm is also based on an exchange algorithm, so it requires a random starting choice design and then considers exchanging every alternative in the starting design with one from a predefined candidate set of all possible alternatives. If an improvement in the design efficiency with respect to the specific criterion occurs in the altered design, then the exchange is performed. The first iteration is completed by finding the best exchange for all alternatives in the starting design. The iteration is continued until no more improvement is possible; see Kessels et al. (2004) for more details of the algorithm. Also, in their study they recommend repeating the algorithm using different starting designs in order to avoid poor local optima. Each starting design is called a try, and the optimal Bayesian choice design is the most efficient design among these tries.

Kessels et al. (2004) compare Bayesian designs generated using the Bayesian modified Fedorov algorithm with those obtained from the RSC algorithm developed by Sándor and Wedel (2001). They show that, based on the underlying design type in their study ($3^4/2/15$), their algorithm provides more efficient designs than those generated using the RSC algorithm. This follows since their design algorithm does not impose any design restrictions such as minimal overlap property on the search algorithm, as opposed to the RSC algorithms developed in Sándor and Wedel (2001). Also, using the MCMF they were able to generate Bayesian optimal designs for the main effects MNL model based on other design criteria (e.g. $\mathcal{A}$, $\mathcal{V}$ and $\mathbf{G}$) rather than restricting themselves to the widely used $\mathcal{D}$-optimality criterion. A comparison between Bayesian designs based on $\mathcal{D}_B$-, $\mathcal{A}_B$-, $\mathcal{V}_B$- and $\mathbf{G}_B$-optimality criteria was then carried

out to investigate their performance in terms of parameter estimation and prediction validity. The comparison showed that $\mathcal{D}_B$- and $\mathcal{A}_B$-optimal designs give more accurate parameter estimates than those based on $\mathcal{V}_B$- and $\mathbf{G}_B$- optimality criteria, while prediction-based criteria such as $\mathcal{V}_B$ and $\mathbf{G}_B$ designs provide more precise predictions. Nevertheless, $\mathcal{D}_B$-optimal designs perform well in term of prediction compared with the prediction criteria.

Kessels et al. (2006), continuing their work started in 2004, have generated a Bayesian optimal design for the main effects MNL model. They used the MCMF based on different design criteria, aiming to compare $\mathcal{D}_B$-, and $\mathcal{A}_B$- design criteria to $\mathcal{V}_B$- and $\mathbf{G}_B$-optimality criteria and their computation time, while emphasising the performance of these criteria in terms of predicative ability. A result from their simulation study shows that $\mathcal{D}_B$-optimal design criterion has the highest expected efficiency in terms of other criteria, particularly $\mathcal{V}_B$- and $\mathbf{G}_B$-design criteria that are developed especially to make precise predictions, and the shortest computation time. They therefore recommend using the $\mathcal{D}_B$-optimality criterion to generate optimal choice design for prediction purpose instead of the prediction criteria, as $\mathcal{V}_B$- and $\mathbf{G}_B$-optimality criteria require a larger computational effort for only a small amount of efficiency gained in predictive ability (Kessels et al., 2006).

The computational time that $\mathcal{V}_B$- and $\mathbf{G}_B$-design criteria consume to generate Bayesian optimal designs using the MCMF algorithm makes them impracticable, particularly for complex design problems where there are many attributes and attribute levels (Kessels et al., 2006). This problem motivated Kessels et al. (2009) to develop another algorithm that could accelerate the computational time of these criteria, to make them more feasible. In their next study, they presented a new algorithm, the adaptive algorithm, that is faster than the MCMF algorithm in generating Bayesian choice designs. There are four features in the adaptive algorithm that contribute to speeding up the calculation of the prediction-based criteria, as below.

1. An economical method is used that calculates only the criterion values of the

design that differ in one profile from another design, through an update approach.

2. Computing the $\mathcal{V}_B$-optimality criterion in an efficient way allows calculating the average prediction variance without need for computing the prediction variance for each profile separately.

3. The use of Meyer and Nachtsheim's (1995) coordinate-exchange algorithm, which changes one coordinate of an alternative at a time, reduces the computational time of the optimal choice designs compared to the MCMF algorithm, particularly for complex large design problems. This is because the MCMF algorithm requires a pre-defined candidate set to replace each alternatives in the starting design with one from the candidate set and then examine the improvement in the design efficiency after each change. This takes a long time for large design problems as the candidate set becomes very large.

4. Minimum potential designs are employed, i.e. a small designed sample of prior parameters is used to compute the Bayesian design criterion. This is considered the main improvement of the new algorithm over the MCMF algorithm.

A full description of the adaptive design algorithm is provided in Kessels et al. (2009). Based on this development in the search design algorithm, the construction of $\mathcal{V}_B$-optimum designs becomes more feasible in practice. However, Kessels et al. (2009) show that constructing $\mathcal{V}_B$-optimum design is faster than one based on a $\mathbf{G}_B$-optimality design criterion. This is because minimising the average prediction variance is faster than minimising the maximum prediction variance over the design region. Recently many studies have been conducted to compare Bayesian designs for different optimality criteria, and the performance of locally and utility-neutral optimum designs to Bayesian designs for the MNL model, as well as the effect of misspecifiying the prior distribution of the preference parameters on the efficiency of the design (e.g. Kessels et al., 2008; Rose et al., 2008; Rose and Bliemer, 2009; Kessels et al., 2011). Kessels et al. (2011b) illustrate that, based on several simulation studies for the underlying design type $2^6/2/8$

constructed using coordinate-exchange algorithm, Bayesian optimal design provides the best estimates for the model parameters compared with the locally and utility-neutral designs. In addition, orthogonal or utility-neutral design produce estimation problems for some simulated data sets, and increases the number of dominant choice sets contained in the design. Also, they state that though Bayesian optimal design might produce a dominant choice set, this can be limited by using an appropriate prior distribution or adapting the design algorithm in such a way that it prevents this type of choice appearing in the design (e.g. using design constraints), which is one of our design considerations.

Software developed specifically to construct Bayesian choice designs is now available, such as the JMP12 software (Kessels, 2010), and the Ngene software based on a syntax programming language (Rose and Bliemer, 2012). In general, the JMP design software is less flexible than the Ngene software, particularly in terms of choice model, prior distribution, design criterion and even the optimisation algorithms available to construct the choice designs in addition to other features. Therefore, the Ngene software is expected to receive more attention in the choice design literature, and particularly in the health economics literature. In this thesis, we investigate the ability of these softwares to construct the Bayesian pairwise choice design for valuing health state utilities while satisfying our design constraints, as will be discussed in Chapter 5.

3.8 Summary

DCEs have been used widely in health economics to value direct and indirect health outcomes. Recently, there has been increased interest in using such techniques to value

health states for provision of the utility values within a QALY framework to replace the cardinal methods such as TTO and SG. However, this technique requires many developments, particularly in terms of experimental design, before it can be established as an alternative to the cardinal methods.

This chapter reviewed the main issues with the experimental design used to construct choice designs in health economics in general, and for estimating health state utilities within the QALY scale in particular, and identified the main requirements in the experimental designs for successful implementation of DCEs to estimate health state utilities. Generally, the review showed that the main design issue is that most of the choice experiments constructed for valuing health state utilities are generated using orthogonal array designs or are based on other required statistical properties such as level balance, and minimal overlap. Restricting the construction of choice designs to these properties might result in dominant and implausible combinations of the attribute levels that reduces the design efficiency. However, imposing constraints on the attribute level combinations together with other health evaluation design requirements (particularly including the death state in the choice design to anchor health state utilities into the QALY scale) and the complexity of nonlinear design problems, require deviation from orthogonal design principles and more advanced design methods to construct an efficient choice design.

There has also been increased interest in using computerised search algorithms to construct the choice design, particularly using optimal design theory. However, the construction of optimal designs for nonlinear models, unlike linear models, depends on the unknown model parameters, and hence requires prior information. In the design literature, there have been two solutions to overcome this problem: assuming either prior point estimates (resulting in what are called locally optimal designs), or prior distributions (that result in Bayesian optimal designs). A description of both design approaches together with the most widely used design criteria to generate the optimal designs, for example the $\mathcal{D}$-optimality criterion, has been presented in this chapter.

Because Bayesian design requires minimising the posterior VCM of particular estimators of interest, and this matrix cannot be derived analytically for nonlinear models, and particularly for DCMs, we also reviewed different methods for approximating this matrix either asymptotically using the likelihood function or exactly using the posterior density.

Although Bayesian designs provide more robust design solutions than local optimal designs, as they account for the uncertainty around the possible parameter values, in health economics less attention has been paid to Bayesian designs. Designs have been mainly based on zero priors for the unknown preference parameters, which results in utility-neutral designs, to simplify design construction. However, this assumption is unrealistic, since it assumes no preference for the attribute levels across alternatives, and might reduce the choice design efficiency. In particular, design have been restricted to the optimal design principles (orthogonality, level balance, minimal overlap and utility balance) defined by Huber and Zwerina (1996) who state that jointly satisfying these principles returns an optimal choice design. Nevertheless, for large and more complex designs that involve real constraints (e.g. avoiding dominant and implausible choices) these principles might conflict with each other, and even satisfying these principles might not produce efficient design as illustrated in Street and Burgess (2007, pp.89-91).

This chapter therefore discussed advanced work in optimal design theory, particularly Bayesian optimal designs. We reviewed the advanced algorithms and software that directly optimise the design criterion and allow for incorporating prior distributional information into the construction of the choice designs, to investigate the possibility of adapting these algorithms to our design constraints (e.g. avoiding dominant and implausible health states) and employing them to construct better experimental designs and obtain more reliable estimates for health state utilities. Now, since Bayesian designs require a prior information about the parameter, in Chapter 4, therefore, we analyse real choice data sets for the underlying classification system (i.e. AQL-5D

system) to derive appropriate prior distributions for the unknown parameters of our choice model.

Chapter 4

Analysis of the AQL-5D Data

4.1 Introduction

In this chapter, we illustrate how preference data can be used to estimate the utilities of health states defined by the AQL-5D classification system, and to which extent the type of data can affect the estimated values. In particular, we analyse two empirical preference data sets, elicited using TTO and DCE techniques, using classical and Bayesian approach to estimate asthma health state utilities on the QALY scale. The asthma health state is described by the AQL-5D classification system described in Section 2.3.2. This descriptive system has five attributes: concern about having asthma, shortness of breath, weather and pollution, the impact of having asthma in sleep and activity limitation. Each attribute has five levels of severity ordered from 'very mild' to 'more severe' as shown in Table 2.1. Of course, such a classification system produces too many health states for direct valuation of each individually. Therefore, a sample of

AQL-5D health states is selected to be evaluated using TTO and DCE methods and then a statistical model can be used to predict the utility value for all possible health states defined by the AQL-5D classification system.

This chapter starts in section 4.2 by describing both data sets and methods used for selecting and valuing health states. Methods used for valuing all possible health states defined by the AQL-5D classification system based on the data, concerning only a small sample of states, is recalled in Section 4.3. In Section 4.4, we fit an ordinary linear model and a logit model for the TTO and DCE data, respectively, to infer the parameters in the identified utility model and estimate the utility values for all possible health states defined by the AQL-5D classification system in a classical manner. A Bayesian approach is then used for the same purpose while providing a simple description of the uncertainty in the utility estimates in Section 4.5. In particular, we present a full Bayesian analysis for the TTO data in order to use the posterior inferences of the parameters as a source of prior distribution in constructing Bayesian optimal design for the same case study (AQL-5D) in later chapters. In addition, we consider reanalysing the DCE data using logit model, though it has been analysed in a Bayesian manner previously using both logit and probit models in Cain (2011), mainly to compare the results in terms of uncertainty in the mean utility values with the ones obtained from the TTO Bayesian model. A summary and conclusion of the analysis are presented in Section 4.6.

4.2 Data Description

In this section, we provide a description for both TTO and DCE valuation surveys used to elicit AQL-5D health state preferences/utilities. Sections 4.2.1 and 4.2.2 describe the TTO and DCE data sets used in this analysis, respectively, as well as the design methods used to select health states presented to respondents.

4.2.1 TTO Data

The TTO data were elicited for a sample of AQL-5D health states based on a representative sample from general UK population in south Yorkshire described in Yang et al. (2007). A total of 307 people were interviewed to elicit the TTO value for 98 health states out of the 3,125 possible ones defined by the AQL-5D classification system. The selection of the states was based on balanced design, which ensured that every level of every attribute had an equal chance of being combined with each level of other attributes. These health states were stratified into mixed severity groups based on the sum of their attribute levels, and then allocated into 14 blocks in which each block had 7 health states in addition to the worst state defined by the classification (state 44444). This ensured that each respondent, to whom allocated one block, was all received a set of states balanced in terms of severity, and each state was valued the same number of times except for the worst health state, which is valued by all respondents. Thus, in total there were 2,456 (307×8) health state utility values generated by all respondents.

As a warm-up exercise to help respondents understand the task of eliciting presence for different health states, respondents were ask to rank 10 states in order of preferences. This ranking task involved the 7 intermediate AQL-5D health states allocated in the

blocks in addition to the best health state (state 00000), the worst health state (state 44444) defined by the classification system and immediate death. Each respondent was then asked to elicit the TTO value as described in Section 2.4.1.1 for a practice health state, that is excluded from the analysis, followed by a valuation for the 7 intermediate states allocated in each respondent's block and the worst health state. The valuation study used time board, that is a visual aid, to elicit the TTO value described by the Measurement and Valuation of Health group. This version of the TTO technique was shown to be more reliable than the one without visual aid (Gudex, 1994).

In addition to the TTO questionnaires, respondents were also asked to report their socio-economic characteristics, general health, and which health service they used, as well as other questions related the difficulties in the preference elicitation method used. Table 4.1 summarises the general characteristics of the interviewed sample, showing their sex, age and experience of having asthma.

Table 4.1: Characteristics of the respondents in the TTO valuation survey

		Number	***Percentage (%)***
Sex	Male	139	45.3
	Female	168	54.7
Age	18-25	34	11.1
	26-35	57	18.6
	36-45	61	19.9
	46-55	50	16.3
	56-65	45	14.7
	> 65	60	19.5
Having asthma	Yes	53	17.3
	No	254	82.7
Total		307	100

These demographic terms and personal characteristics have been considered in the classical modelling of the TTO data in Yang et al. (2007) as illustrated in Section 4.4.1. Nevertheless, they showed that model excluding these terms perform better in terms of the predictive ability of the mean utility values. Therefore, none of these personal

characteristics are considered in our classical or Bayesian model for the TTO data.

4.2.2 Discrete Choice Data

In this section, we present another valuation study that used the DCE technique to estimate the AQL-5D health state utilities. In the valuation survey, the choice data were collected for 32 pairwise comparisons from 307 people who had been interviewed in the TTO valuation survey and consented to the postal DCE survey as described in Brazier et al. (2009).

As mentioned earlier, the large number of health states produced by the classification system make it infeasible to value all possible pairwise comparisons, that is $\binom{3125}{2} = 4,881,250$. An experimental design is needed to reduce the number of possible pairwise comparisons to a manageable number of health state comparisons, while still being able to infer the valuation for all possible health states defined by the classification system.

Thus, in this valuation survey, pairs were selected for the choice design based on statistical properties: level balance, orthogonality, minimal overlap and utility balance, using an application in the statistical package SAS developed by Huber and Zwerina (1996). The programme produced 24 pairwise comparisons from the AQL-5D, and these were allocated randomly into four versions of a questionnaire, each with six pairwise choices. Two additional pairwise comparisons, which compare death state to the worst health state defined by the classification system (AQL-5D health state 44444) and another health state in the AQL-5D that could be considered worse than death by some respondents (AQL-5D health state 33244), were added to all versions of the questionnaire to be valued by all respondents, as shown in the Appendix A.2.1. We call this design a level balanced design (LBD), for ease of reference in later chapters

Respondents were asked to value a practice pairwise comparison before starting the actual valuation survey, to familiarise themselves with the choice task, which is excluded

from the final discrete choice data. They were then asked to state which health state they preferred for the six pairs of states in addition to the two pairs of death state. Out of 307, potential respondents, 168 people returned a completed questionnaire with only eight observations missing for some choice tasks, generating 1,336 observed pairwise comparisons. Table 4.2 shows the characteristics of the respondents who participated in the DCE survey.

Table 4.2: Characteristics of the respondents in the postal DCE valuation survey

		Number	***Percentage*** (%)
Sex	Male	72	43
	Female	96	57
Age	18-25	6	3.6
	26-35	22	13.1
	36-45	28	16.7
	46-55	38	22.6
	56-65	39	23.2
	> 65	35	20.8
Having asthma	Yes	44	26
	No	110	66
	Unknown	14	8
Total		168	100

As in the TTO volution exercise, most participants were from the general public though the AQL-5D is a condition specific classification system. This is because health decision makers such as those at NICE recommended using general public values rather than collecting the data from the asthma patients only. Nevertheless, all participants should understand the condition and complete the questionnaire while considering themselves as having asthma to provide more reliable preference data.

4.3 Method for Modelling Health State Utilities

In this section, we recall the method for generating utility values for all health states defined by a classification system, particularly for the AQL-5D classification system, using a statistical model. We mentioned that the AQL-5D classification system produces too many health states, and it is impractical to elicit health state utility values for all possible health states defined by the classification system. Therefore, a sample of these health states were evaluated using TTO and DCE methods. Nevertheless, we wish to obtain the health state utility for all AQL-5D health states. To achieve this, a parametric model is fitted to the TTO and discrete choice data, parameters in this model are evaluated, and then utilities for any state defined by the AQL-5D classification system are inferred.

In Chapter 2, we defined health state utility as function of the attribute levels that make up the health states, giving the population mean utility $g(\boldsymbol{x}_{ij})$, plus a random component that describes the variation around the mean as presented in equation 2.3.1:

$$U_{ij} = g(\boldsymbol{x}_{ij}) + \epsilon_{ij}.$$

In general, the mean health state utility is defined as a linear additive function of the attribute levels of the classification system as

$$g(\boldsymbol{x}_{ij}) = 1 - \boldsymbol{\beta}\boldsymbol{x}_{ij}^{T},$$

where $\boldsymbol{x}_{ij}$ is a vector of 20 dummy variables that defines the AQL-5D health states with elements defined as

$$x_{\lambda\delta} = \begin{cases} 1 & \text{if attribute } \delta \text{ of health state } \boldsymbol{x}_{ij} \text{ is at level } \lambda \text{ or higher,} \\ 0 & \text{otherwise.} \end{cases}$$

For the purpose of anchoring the utility values produced by the DCE data, we added a dummy variable that defines the death state. This dummy variable is equal to zero unless the health state represents the death state, in which case the first 20 elements are 0 and the last element is 1.

The vector $\boldsymbol{\beta}$ consists of 21 unknown parameters where the first 20 elements represent the decrease in utility associated with moving one level on one attribute hence they are expected to be positive, and the last one corresponds to moving from perfect health to immediate death, which is set at 1 for scaling purposes, as described in Section 2.3.3. This statistical model allows us to estimate the utility values for any state defined by the AQL-5D classification system within the required scaled of the QALYs (0-1 scale), while retaining the possibility of having negative values for health states worse than death.

To be able to estimate the health state utility for all possible states defined by the classification system, we have to make inferences about the preference parameter values, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{20})$. Therefore, in the following sections, we illustrate how TTO and choice data are modelled to infer the parameter vector $\boldsymbol{\beta}$ first in a classical framework in Section 4.4, and then using a Bayesian approach in Section 4.5.

4.4 Classical Inference for Health State Utilities

In this section, we describe the structure of the TTO and DCE models and summarise their classical inferences using the maximum likelihood approach as illustrated in Sections 4.4.1 and 4.4.2, respectively.

4.4.1 Modelling TTO Data

Yang et al. (2007) present an analysis of TTO data elicited for the 98 AQL-5D health states and the worst health state (AQL-5D health state 44444) using different regression models. In their study, one-way error components random effect and fixed effect models were considered to model the TTO data at individual and aggregate levels. The general form of this model is

$$U_{ij} = g(\boldsymbol{x}_{ij}\boldsymbol{\beta}^T + \boldsymbol{r}_{ij}\boldsymbol{\theta}^T + \boldsymbol{z}_i\boldsymbol{\delta}^T) + \epsilon_{ij}, \tag{4.4.1}$$

where U_{ij} represent the TTO value for health state $\boldsymbol{x}_{ij}$ evaluated by respondent i; $\boldsymbol{x}_{ij}$ is a vector of 20 dummy variables, $x_{\lambda\delta}$, for each level λ of attribute δ of the classification system, where $\lambda = 0$ is taken as the baseline level for each attribute; $\boldsymbol{r}$ is a vector of interactions between attributes and $\boldsymbol{z}$ is a vector of personal characteristics such as sex, age and asthma condition; ϵ_{ij} is an error term whose autocorrelation structure and distributional properties depend on the assumptions underlying the model used. The parameter vectors $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ are vectors of the unknown model parameters associated

with attribute levels of the AQL-5D, the interaction between the attributes and the personal characteristics, respectively.

All proposed models were estimated using 1-TTO values as the dependent variable, where 1 represents the utility of full health (AQL-5D health state 00000) in order to avoid a negative values for the preference parameters $\boldsymbol{\beta}$. This variable indicates how far a given health state is away from full health, i.e. the utility lost when moving from perfect health to health state $\boldsymbol{x}_{ij}$. The models were then compared on the basis of different criteria, such as adjusted R square, goodness of fit, likelihood ratio, the size and the significance of the parameter estimates, and the over all predictive ability. However, since predicting utility values for all health states is the main aim of modelling preference data, the predictive ability was chosen as the main criterion to compare and select the best model.

In this way, Yang et al. (2007) select both the ordinary linear model at the individual level and at the aggregate level with no interactions or demographic terms (e.g. age, gender) as being the most appropriate. Further, in economic evaluation the average utility value of any health state is required rather than the individual health state utility; therefore the model at aggregate level is regarded as the best model for estimating health state utilities.

Following the results in Yang et al. (2007), we re-analyse the TTO data at the aggregate level using the simple regression model, with no interactions or demographic terms, and summarise the main results and the estimated mean utilities. We then investigate the uncertainty around the utility estimates using a Bayesian approach, and compare it with that estimated using maximum likelihood estimators.

The simple regression model takes the form

$$1 - U_j = \boldsymbol{x}_j \boldsymbol{\beta}^T + \epsilon_j, \quad j = 1, \ldots, 99, \tag{4.4.2}$$

where $1 - U_j$ is the observed mean 1-TTO value for health state $\boldsymbol{x}_j$, $\boldsymbol{x}_j$ and $\boldsymbol{\beta}$ are,

as defined previously, vectors of dummy variables defines the underlying health state and the unknown parameters, respectively, and ϵ_j is the error term associated with the mean utility value.

In matrix notation, we can write this as follow

$$1 - \boldsymbol{U} = \boldsymbol{X}\boldsymbol{\beta}^T + \boldsymbol{\epsilon}, \tag{4.4.3}$$

where $\boldsymbol{U} = (U_1, \ldots, U_J)^T$ is a column vector of J observed TTO values, $\boldsymbol{X}$ is $J \times 20$ design matrix with rows each represents one health state; $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{20})$ is a vector of the unknown model parameters, and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_J)^T$ is a column vector of random errors assumed to be independent and normally distributed with mean zero and equal variance, i.e. $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.

The estimation of the parameter values can be derived using the maximum likelihood given that $1 - \boldsymbol{U} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}^T, \sigma^2\mathbf{I})$. Nevertheless, under the normal assumption of the random errors, the maximum likelihood estimators (m.l.e) are equivalent to the ordinary least square(OLS) estimators. The ordinary least squares regression coefficients are derived by minimising the difference between observed and fitted values, i.e. minimising the residual sum of square. The general form for the OLS estimators is given by

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(1 - \boldsymbol{U}), \tag{4.4.4}$$

To estimate the range of each parameters values and assess the model fit, one need to estimate the scale parameter of the error variance, σ, since the variance-covariance matrix of the least square estimators depends on the value of σ as

$$Var(\hat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1} \tag{4.4.5}$$

Using the least squares estimators, an estimate for σ is computed using the residual

sum of squares as

$$\hat{\sigma} = \frac{\sum_{j=1}^{J}[(1-\widehat{U}_j)-(1-U_j)]^2}{J-k}, \tag{4.4.6}$$

where $1-\widehat{U}_j = \boldsymbol{x}_j\hat{\boldsymbol{\beta}}^T$ is the estimated loss in the mean utility value for state $\boldsymbol{x}_j$, J and k are the total number of observation and unknown model parameters, respectively.

Table 4.3 illustrates the parameter estimates after fitting the OLS model to the 1-TTO data. Each parameter estimate represents the incremental decrease in mean utility when moving one level on one attribute. Therefore, the estimated decrease in mean utility from the best level to the worst level of an attribute is computed as the sum of these increments. For instance, the estimated mean decrease in utility when moving from level 0 on the attribute concern (no concern about having asthma at all) to the most severe level 4 (feel concern about having asthma all of the time) is $0.02989 + 0.02624 + 0.00709 + 0.00265 = 0.06587$.

Table 4.3: OLS estimators for TTO model

	Attribute				
Level	Concern	Breath	Weather	Sleep	Activities
1	0.02989	-0.00845	0.00589	0.04550	0.01244
2	0.02624	0.03000	0.01917	0.01482	0.05134
3	0.00709	0.07615	0.02399	0.01927	0.12366
4	0.00265	0.00090	0.05741	0.02265	-0.00654
σ			0.07325		

The size of the parameter estimates ranges from -0.00845 to 0.12366, and there is no clear pattern in these sizes. Nevertheless, for most attributes, the decrease in mean utility when changing from level 2 to level 3 is larger than changing from level 3 to 4. This seems consistent with the description of these attribute levels, as the state of suffering from having asthma some of the time (level 2) is closer to a little of the time (level 1) than those suffering from asthma most of the time (level 3), and similarly the description of level 3 (most of the time) is closer to level 4 (all the time). For instance, considering the attribute breath in the following AQL-5D states, 13112, 14112 and

12112, a person would have a similar mean utility for the state suffering from short of breath most of the time as a result of having asthma (13112) and that suffering from the same symptom all of the time (14112) than to state 12112. The largest decrease in mean utility is a move from level 2 to level 3 for the attribute activities in which the mean utility is decreased by 0.12366.

The TTO model produces the expected positive sign for all the estimated parameters, except for level 1 and level 4 for the attributes breath and activities, respectively. This is because no constraints were imposed in the parameter estimates as in Yang et al. (2007), which results in a negative value for the levels of these attributes. These negative values are inconsistent with the natural ordering of the attribute levels of the AQL-5D, but statistically insignificant (i.e. these attribute levels do not influence respondents' valuations and, hence, do not differ from zero).

Using the OLS estimates for the preference parameters, we can now calculate the estimated mean utility value for the 99 health states presented in the TTO survey or any health state defined by the AQL-5D system as

$$\widehat{U}_j = 1 - \boldsymbol{x}_j \hat{\boldsymbol{\beta}}^T. \tag{4.4.7}$$

Figure 4.1 illustrates the estimated mean utility values, i.e. the TTO values of each health state averaged over all respondents, for 99 AQL-5D health states presented in the TTO survey together with the mean observed TTO value. Also, it shows the 95% confidence interval of the estimated mean health state utilities, $\widehat{U}_j$, where the confidence interval of the mean utility of each health state is computed as

$$\widehat{U}_j \pm t_{J-k,0.975}\, \boldsymbol{x}_j \, \frac{\hat{\sigma}\mathbf{I}}{\sqrt{J}} \, \boldsymbol{x}_j^T. \tag{4.4.8}$$

The points representing the mean TTO value of each state, as well as the upper and lower bound of the confidence intervals, are joined in this and subsequent plots simply to

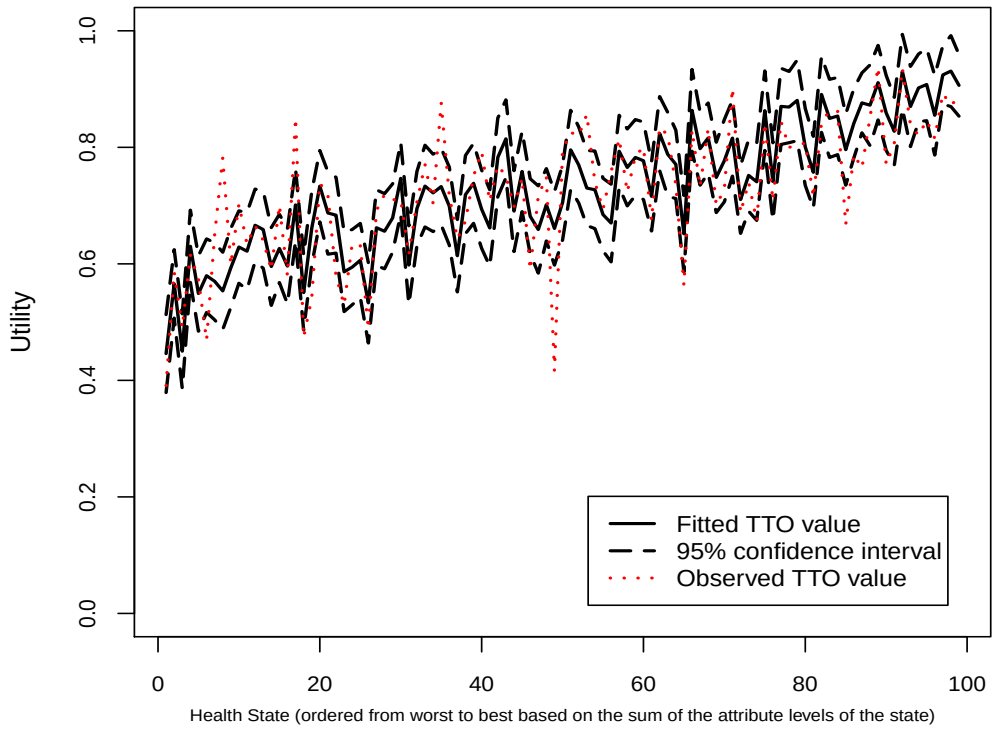

Figure 4.1: the estimated mean utilities for the 99 AQL-5D health states evaluated in the TTO survey together their 95% confidence intervals and the observed TTO values

aid visual interpretation. In general, the model fits the data reasonably well with a little over-estimation for health states near the best state (state 11111). Nevertheless, most of the observed mean utility lies within the 95% confidence interval of the fitted utility values, but this does not account for the uncertainty in the estimated mean utility values, and consequently the QALY values required for cost-effectiveness analysis.

We mentioned earlier in Section 4.3 that the mean utility values should be lie within the QALY scale where full health has a utility of one and the death state has utility of zero, while allowing health states worse than death to take negative values. Therefore, to find the complete range of the mean utility, it is important to compute the mean utility value for the worst health state (AQL-5D health state 44444) and compare it

with the utility value of death (zero). The mean utility of this state is represented by 1 minus the sum of all the preference parameter estimates shown in Table 4.3 which is estimated as 0.4459, with a 95% confidence interval that ranges from 0.37899 to 0.51280. This indicates that this health state is considered to be substantially better than death. This might indicates that there is a substantial number of poorer health states that are not described by the AQL-5D system, or most respondents are not willing to trade much time in order to live healthier and most of them are considering all health states as worth living.

4.4.2 Modelling Discrete Choice Data

In Chapter 2, we mentioned that health state utilities cannot be elicited directly from a DCE task, and therefore choice models are used to relate the observed choices to the identified utility function such that it infers the parameters of the utility model. The pairwise choice data presented in Section 4.2.2, for the AQL-5D, was analysed by fitting a probit model to the data in Brazier et al. (2009), and in Cain (2011) using the logit and probit models.

Assuming that the AQL-5D choice data are independent over respondents, and the random errors are independent over time in repeated choice task, then the logit model would be an appropriate model to capture the dynamics of repeated choice data. Cain (2011) showed that both logit and probit models with main effects are appropriate for the AQL-5D discrete choice data. Other models such as the mixed logit model, that allows for preference heterogeneity across individuals, might be more appropriate to describe human choice behaviour. However, in this thesis, we consider the logit model with main effects only and no interaction terms, to simplify the choice design problem as it has a close form for the choice probability and, hence, the likelihood function. Therefore, in this section, we re-analyse the DCE data for the AQL-5D to (1) describe our main design model, i.e. the logit model with main effects only, and (2) compare the DCE results to those obtained from the TTO model particulary in terms of parameters

and utility estimates.

In a pairwise choice experiment, each individual is asked to select the preferred health state in each pair $C_s = \{\boldsymbol{x}_{i1s}, \boldsymbol{x}_{i2s}\}$. The probability that individual i selects the first health state, $\boldsymbol{x}_{i1s}$, is equivalent to the probability that the utility individual has for health state $\boldsymbol{x}_{i1s}$ is greater than the utility for the second health state $\boldsymbol{x}_{i2s}$. That is:

$$P_{i1s} = P[g(\boldsymbol{x}_{i2s}) + \epsilon_{i2s} < g(\boldsymbol{x}_{i1s}) + \epsilon_{i1s}], \tag{4.4.9}$$

or if the second alternative is the death state,

$$P_{i1s} = P[0 < g(\boldsymbol{x}_{i1s}) + \epsilon_{i1s}], \tag{4.4.10}$$

where $g(\boldsymbol{x}_{ijs})$ is the mean health state utility defined as in Section 4.3.

In Section 2.5.3, we showed that if the random errors in the utility model, ϵ_{ijs}, are assumed to be independent and identically type 1 extreme value distributed, then the probability of choosing a health state is given by the multinomial logit model defined in Equations (2.5.12) and (2.5.14). From these equations, it can be deduced that the probability that individual i chooses the first alternative, $\boldsymbol{x}_{i1s}$, is

$$P_{i1s} = \frac{1}{1 + \exp\left(\frac{g(\boldsymbol{x}_{i2s}) - g(\boldsymbol{x}_{i1s})}{\sigma}\right)}, \tag{4.4.11}$$

and for death state comparisons

$$P_{i1s} = 1 - \exp\left[-\exp\left(\frac{g(\boldsymbol{x}_{i1s}) + \mu}{\sigma}\right)\right], \tag{4.4.12}$$

where the death state is the second alternative, $\mu = -0.5772\sigma$ and σ are the location and the scale parameters of the random errors ϵ_{ijs}, respectively.

Each choice made for particular pair by a specific respondent can be considered as an independent draw from a Bernoulli distribution, in which it has a value of 1 if the

first health state is chosen and zero otherwise. This follows since random errors are assumed to be independent over alternatives and respondents. Therefore, we derive the likelihood function of the logit model for all the observed choices, made over all pairs, based on this assumption as

$$L(\boldsymbol{y}|\boldsymbol{\beta},\sigma) = \prod_{s=1}^{32}\prod_{i=1}^{N_s} P_{is}^{y_{is}}(1-P_{is})^{(1-y_{is})}, \tag{4.4.13}$$

where

$$y_{is} = \begin{cases} 1, & \text{if the first state } \boldsymbol{x}_{i1s} \text{ is chosen;} \\ 0, & \text{otherwise.} \end{cases} \tag{4.4.14}$$

and P_{is} is the corresponding choice probability for the first health state in each pairs s, where for $s = 1, \ldots, 24$, non-death comparisons, the choice probability is defined as in Equation (4.4.11), and for $s = 25, \ldots, 32$, death comparisons, the choice probability is defined as in Equation (4.4.12); and N_s is total number of respondents evaluating choice set s where $\sum_{s=1}^{32} N_s = N = 168$.

The log likelihood function is derived then as

$$l(\boldsymbol{y}|\boldsymbol{\beta},\sigma) \;=\; \sum_{s=1}^{32}\sum_{i=1}^{N_s} y_{is}\log(P_{is}) + (1-y_{is})\log(1-P_{is}) \tag{4.4.15}$$

The preference parameters identified in the utility model can then be estimated by maximising the likelihood function. Nevertheless, the likelihood function defined in Equation (4.4.15) has many parameters and it is non-linear. Therefore, it is difficult to obtain the maximum likelihood estimators (m.l.e) analytically and, hence, a numerical method should be sought. Here, the maximum likelihood estimators are obtained numerically using optimisation algorithms in R software called 'optimisation'. The algorithm searches for the optimal values of the parameters that maximise the log likelihood function over a given parameter space. The resulting parameter estimates are illustrated in Table 4.4.

Similar to the TTO model, the parameter estimates presented in Table 4.4 show

Table 4.4: Maximum likelihood estimators for logit model

	Attribute				
Level	Concern	Breath	Weather	Sleep	Activities
1	0.00362	0.00412	$7.070e^{-07}$	$4.451e^{-06}$	0.03164
2	$4.391e^{-07}$	$2.645e^{-08}$	0.02173	0.05593	0.02682
3	0.13354	0.12490	0.07937	0.02225	0.24128
4	0.03375	0.00001	0.02555	0.02927	0.00017
σ			0.24140		

the incremental decrease in the mean utility value when changing an attribute by one level, after normalising using the death coefficient to anchor the utility value into the QALY scale. The preference parameter estimates range from $2.645e^{-08}$ to 0.24128, where the smallest change in mean utility is approximately zero when rounded to five decimal places. This change is associated with moving from level 1 to level 2 for the attribute breath. This means that health states that differ only on this attribute level would have similar utility value, for instance, states 12112 and 11112. There are several other parameters that also have very small values. For example, a change from level 0 to level 1 in the weather and sleep attributes does not produce a substantial change in the estimated mean utility from the full health state (AQL-5D health state 00000). The largest decrease in mean utility is associated with moving from level 2 to level 3 for the activities attribute which is consistent with the result for the TTO model.

Imposing constraints on the estimation procedure produces positive coefficients for all preference parameters of the fitted logit model. This produces an incremental decrease in mean utility that is consistent with the logical order of the attribute levels of the AQL-5D, i.e. the decrease in mean utility increases as we move to more severe levels in each attribute.

Using the parameter estimates from the logit model, shown in Table 4.4, we can estimate the mean utilities for all the health states presented in the DCE survey or

any health state defined by the AQL-5D system as

$$g(\boldsymbol{x}_{ij}) = 1 - \hat{\boldsymbol{\beta}}\boldsymbol{x}_{ij}^{T}.$$

Figure 4.2 shows the estimated mean health state utility values for these states, where health states in this plot, and the subsequent plots of the mean health state utilities, are ordered from worst to best state based on the sum of the attribute levels of the state, where larger value of this sum refers to more worse health state.

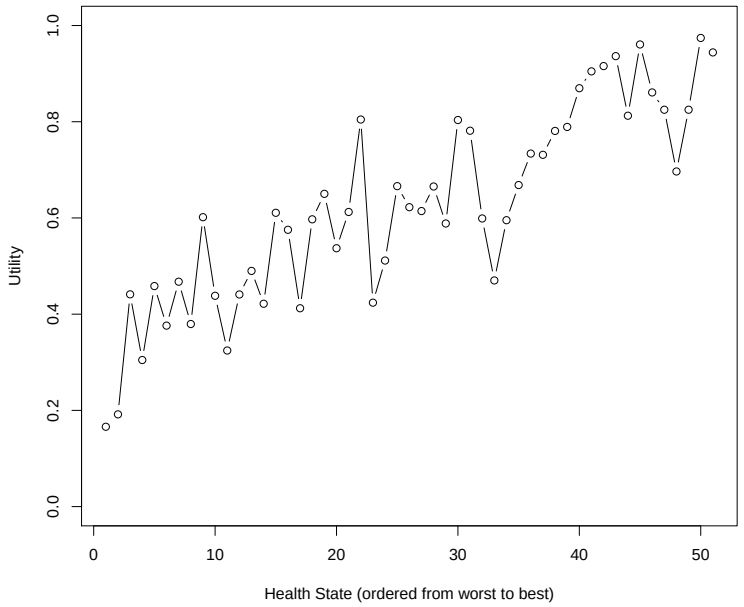

Figure 4.2: The estimated mean utilities for the 51 AQL-5D health states evaluated in the DCE survey using the maximum likelihood estimators of the preference parameters in the logit model

We notice that the mean health state utility of the worst health state is estimated as 0.16602, and so it is considered to be better than death state. Thus, the estimated utility values for all AQL-5D health states using the fitted logit model are ranged between 0.16602 and 1, with no health state being worse than death.

Since we cannot observe the utility values directly from the DCE task, then we

investigate the model's predictive ability by plotting the predicted choice probabilities for the first health state in each choice set against the observed choice probabilities. Figure 4.3 presents the fitted choice probability of choosing the first health states against the observed choice probability in the choice survey together with the 95% confidence interval of the fitted proportions. The confidence interval is computed using the normal approximation of the binomial proportion as

$$\hat{P}_{is} \pm Z_{1-\alpha/2}\sqrt{\hat{P}_{is}(1-\hat{P}_{is})/N_s},$$

where $\hat{P}_{is}$ is the fitted proportion and, $Z_{1-\alpha/2} = 1.96$ is the percentile of the standard normal distribution, and N_s is total number of respondents evaluated choice set s.

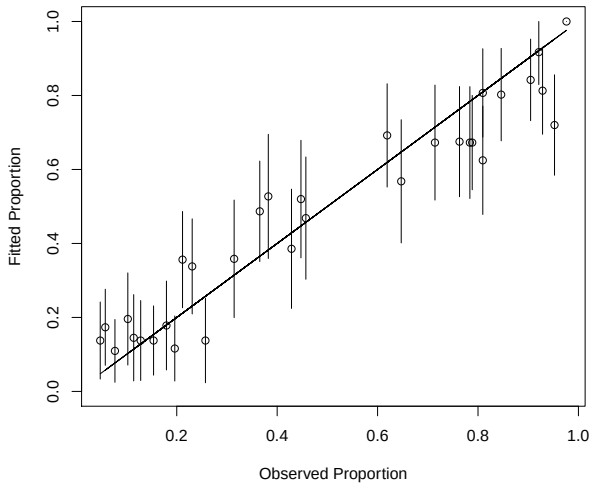

Figure 4.3: The estimated logit choice probability of the first health state in each pairwise comparison together with its 95% confidence interval against the observed choice probability in the DCE survey

In general, we might conclude that the logit model fits the data reasonably well, though there is a slight suggestion of over-estimations when the proportion is low and under-estimation for larger proportions. Nevertheless, most of the observed proportion lie within the 95% confidence interval of the fitted proportion except for some points.

4.4.3 Comparing Classical Inference for Health State Utilities

In this section, we compare the classical inference for both TTO and logit models. We should not expect to obtain the same results for the TTO and DCE models (logit model), since they follow different valuation procedures and completely different model assumptions. However, we could compare the models in terms of the size of the coefficients as well as the predictive ability of the model. In terms of the prediction ability of the model, we compare the pattern of the models' predictions to the observed TTO values, though it is not necessarily the case that the TTO values represent the correct valuations for a given health state as it is affected by other non-health factors, such as time preference (Brazier et al., 2007).

From Tables 4.3 and 4.4, it can be seen that logit model produces smaller parameter estimates than the TTO model, but larger values for those parameter associated with the decrease in mean utility when moving from level 2 to level 3 in all attributes. However, the models have similar patterns for the incremental decrease in mean utility. For instance, in both models the largest decrease in the mean utility is associated with a change from level 2 to level 3 for the attribute activities, though it is larger for the logit model.

The models' predictions for the observed mean TTO values are illustrated in Figure 4.4. The TTO model predicts the observed mean TTO values more accurately than the DCE model, as expected, with mean absolute differences from the observed TTO values of 0.0472 and 0.106, for the TTO and the logit predictions respectively. However, this might be related to the fact that the numbers of respondents and health states evaluated in the DCE survey are smaller than the ones used for the TTO exercise as shown in Table 4.5.

In general, the DCE model produces prediction values that are less than the observed TTO values for more severe health states (involving a low TTO value), whereas for mild AQL-5D health states the DCE predictions are more evenly spread around

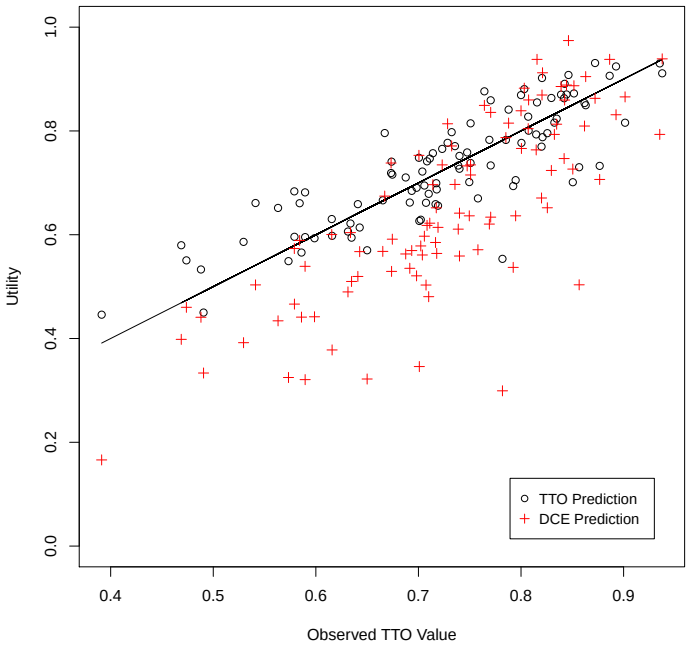

Figure 4.4: A plot of the predicted mean TTO values for the 99 AQL-5D health states presented in the TTO survey using TTO and logit models against the observed mean TTO values

with the DCE model producing similar predicted values to the TTO model for the mild health states. The estimated mean utility for the worst health state, using both models, is greater than zero, which indicates that generally this health state is considered better than death. However, the predicted value for the worst health state from the DCE model is nearer to zero (0.16602) than the equivalent value from the TTO model (0.4459). This means that there are some more worse asthma health state that cannot be described by the attributes and the levels in the AQL-5D classification system, since the utility value of the worst health state is far apart from the utility of death (i.e. the worst health state is not comparable to death state). Nevertheless, this might be related to the difficulty of the TTO valuation task and the effect of the time in respondents' preferences as discussed in Section 2.4.2 rather than the description of the classification system itself, as most respondents not willing to trade much of their life expectancy off against being more healthy though with the worst health

Table 4.5: The total number of respondents, health states, observations as well as the method used to select the health states in the TTO and DCE surveys

	Data	
	TTO	DCE
Respondent	307	168
Health state	99	52
Observation	2456	1336
Design	balanced design	Huber and Zwerina (1996) design

state (44444) and most of them considered this state worth living compared to death. Whereas, simpler evaluation task such as the pairwise comparison eliminates this effect and force respondents to trade-off between this state and the death state.

However, neither form of classical model estimates delivers an easily interpreted measure of the uncertainty in the parameter estimates, and hence the estimated mean utility values. Although standard errors are, at least approximately, provided for each estimated β, uncertainty is much more easily propagated and handled within a Bayesian framework where it can be represented by a probability distribution rather than range of values that represented by the confidence intervals. Also, health economists usually require a a probability distribution for the QALY gained (utility gained) instead of a range of values for the QALYs (utilities), and Bayesian approach provides probability distributions of the parameters that allows to account for the uncertainty in the utility values, and, hence, the possible change in the QALY values for being in particular health state. This accounts for the uncertainty in the decision made through the cost-effectiveness analysis by providing a probability distribution for a treatment being cost effective at a particular threshold (e.g. $£20,000 - £30,000$ per QALY gained) when comparing different treatments.

Therefore, in the following section we analyse both data sets using Bayesian methods. This allows us to (1) make a comparison of classical and Bayesian approaches and account for the uncertainty in the utility values, (2) use the probability distributions

of the parameters as prior information for construction of Bayesian choice designs in later chapters.

4.5 Bayesian Inference for Health State Utility

Health state utility values are used to compute the QALYs in an economic evaluation models, and it is important for the decision makers to account for the uncertainty in these quantities in order to arrive at an appropriate decision. However, classical analysis is usually used to give a single point estimate value for the mean utility for each AQL-5D health state, which ignores the uncertainty in the utilities, and therefore Bayesian inference is usually required. In a Bayesian approach, a posterior distribution for each preference parameter is obtained, and consequently a posterior distribution instead of the single value can be obtained for the utility value of each health state defined by the classification system, as well as for the possible change in the QALY values.

Generally, in the Bayesian approach, the prior beliefs $\pi(\boldsymbol{\theta})$ about the unknown parameter vector $\boldsymbol{\theta}$ are converted into posterior beliefs, which represents the probability distribution of the unknown parameters conditional on the observed data, through Bayes Theorem

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{L(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{P_Y(\boldsymbol{y})}, \tag{4.5.1}$$

where $P_Y(\boldsymbol{y})$ is the marginal distribution of the observed data $\boldsymbol{y}$.

Thus our first task is to identify an appropriate prior distribution for the unknown model's parameters.

4.5.1 Prior Distribution

Deriving a posterior distribution for the population mean utility for each health state in the AQL-5D requires a prior distribution for the unknown parameters identified in the utility model. Therefore, a prior is needed for the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)$, where $\boldsymbol{\beta}$ and σ are a vector of the preference parameters and the scale parameter of the random component of the utility function described in Section 4.3.

Firstly we consider selecting an informative prior distribution for the preference parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{20})$. Each parameter $\beta_1, \ldots, \beta_{20}$ represents the decrease in the mean utility when changing one attribute by one level as illustrated in Section 4.3. Therefore, the values of these parameters should be positive to be consistent with the logical order of the attribute level in the AQL-5D. Also, the size of these parameters is not likely to exceed one, since worsening one attribute by one level is not expected to produce a change in utility greater than the change from perfect health to death that is represented by one.

In this analysis, we consider a collection of prior distributions that exhibit these features in the parameter values. First, we propose independent Gamma priors for each element of the parameter vector $\boldsymbol{\beta}$ as in Cain (2011). The probability density function (pdf) of the Gamma distribution is given by

$$f(\beta_i) = \frac{b^a}{(a-1)!}\beta_i^{a-1}e^{-b\beta_i}, \quad i = 1, \ldots, 20, \tag{4.5.2}$$

where a and b are shape and rate parameters, respectively. The mean and variance of the Gamma distribution are given by $\frac{a}{b}$ and $\frac{a}{b^2}$, respectively. Since the β_i values should be positive and are not likely to be greater than 1, we consider different Gamma distributions that fall within the range of [0,1].

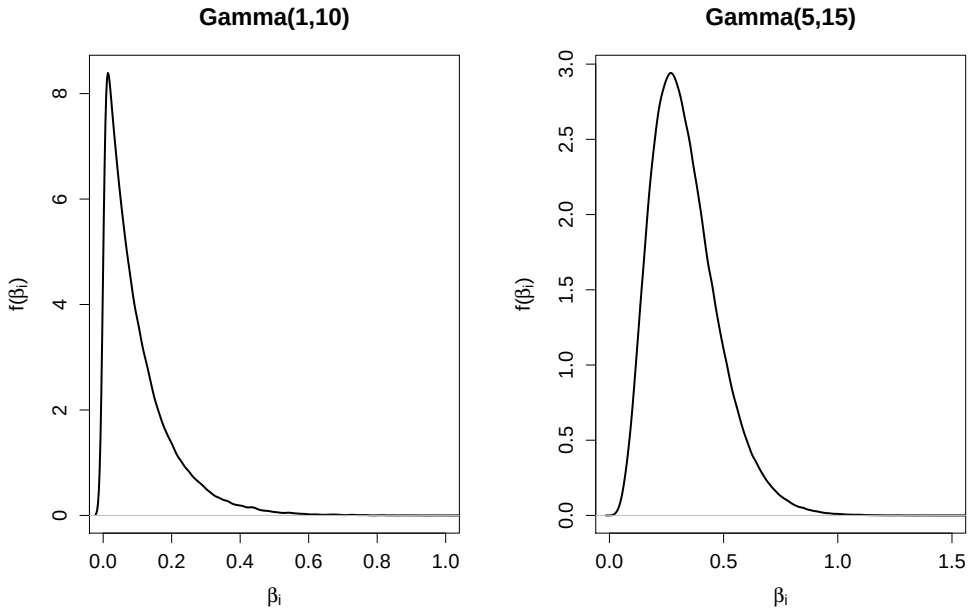

Figure 4.5: A plot of the pdf of Gamma distributions

Three prior distributions, as in Cain (2011), are considered. Firstly Gamma(1,10) and Gamma(5,15) are assumed. These priors give parameter values that are more likely to be close to zero and have a small probability of being greater than 1, as shown in Figure 4.5. The prior distribution with shape parameter $a = 1$ and scale parameter $b = 10$ provides more parameter values that are close to zero than the Gamma(5,15) distribution. Also, it has a small probability for the parameters to be larger than 0.2. Hence, the Gamma(1,10) prior might provide more appropriate prior for the parameters. Nevertheless, we perform the analysis for both priors as we are not sure about the exact range of each parameter value, and to investigate the sensitivity of the posterior inferences to the prior distribution chosen. A continuous uniform distribution between 0 and 1 is considered as well to investigate the sensitivity of posterior inference to the prior distribution. This prior illustrates the extreme case where the prior information about individuals' preferences for attribute levels, β_i, has a substantial amount of uncertainty. The prior assumes that all parameters are equally likely to take any value between 0 and 1. Thus, this can be considered as a vague prior

for the parameter vector $\boldsymbol{\beta}$. In addition, we consider the beta distribution with shape parameters of 1 and 10, respectively, to study the effect of this prior on the posterior inference.

Also, we must consider including a prior information about the scale parameter σ. This parameter should have similar properties as $\boldsymbol{\beta}$, that is the value of this parameter should be also positive and small in magnitude (e.g. not larger than one). This is because the variability in the utility values obtained from different individuals for the same health state is not likely to exceed one. Therefore, for simplicity, we use the same suggested prior distribution for the preference parameters though different distribution can be used.

4.5.2 Obtaining the Posterior Distribution

The posterior distribution is derived as in Equation (4.5.1). However, given the likelihood function of the TTO model, $\mathcal{N}(\mathbf{X}\boldsymbol{\beta}^T, \sigma^2\mathbf{I})$, and the likelihood function of the logit model defined in Equation (4.4.13) together with the prior distributions considered earlier, the posterior distribution for each parameter cannot be derived analytically because the analysis involves non-conjugate priors and complicated integrals. Therefore, Markov Chain Monte Carlo (MCMC) sampling is used to simulate the posterior distribution of the unknown parameter vector, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)$, in the TTO and logit models.

In general, the MCMC method draws values for the parameter vector from the starting distribution (prior distribution), and then updates these draws based on a specific iteration procedure to improve the approximation of the target posterior distribution. Therefore, the posterior distributions of the 21 unknown parameters in the TTO and logit models, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{21})$ where $\theta_i = \beta_i$ for $i = 1, \ldots, 20$ and θ_{21} corresponds to the scale parameter σ, are approximated using MCMC method by drawing several independent sequences for the parameter vector, $\boldsymbol{\theta}^t, t = 1, 2, 3, \ldots$, where the first sequence, $\boldsymbol{\theta}^1$, is generated from the transition distribution $T_t(\boldsymbol{\theta}^t|\boldsymbol{\theta}^{t-1})$ using a stat-

ing values $\boldsymbol{\theta}^0$, and then each sequence $\boldsymbol{\theta}^t$ is drawn from the same distribution based on the previous sample $\boldsymbol{\theta}^{t-1}$ (Gelman et al., 2004, pp.285-287).

The first sequence vector is generated by updating an initial vector $\boldsymbol{\theta}^0$ drawn from the starting distribution of the parameters $\pi(\boldsymbol{\theta})$. At any time t, where $t = 1, 2, \ldots,$ the state of the Markov chain at time t, $\boldsymbol{\theta}^t = (\theta_1^t, \ldots, \theta_{21}^t)$, is updated to the state at time $t+1$, $\boldsymbol{\theta}^{t+1} = (\theta_1^{t+1}, \ldots, \theta_{21}^{t+1})$, using the single component Metropolis-Hastings algorithm. The algorithm updates the state of the chain by updating one element of $\boldsymbol{\theta}$ at a time, that is in 21 steps taken in natural order. Each complete update represents one iteration of the MCMC. To illustrate this further denote $\boldsymbol{\theta}_{-i}$ to be the vector of all elements of $\boldsymbol{\theta}$ except θ_i

$$\boldsymbol{\theta}_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots \theta_{21}),$$

and define $\boldsymbol{\theta}^t_{-i}$ as the state of $\boldsymbol{\theta}_{-i}$ after updating the $i-1$ the component of $\boldsymbol{\theta}^t$ at time $t+1$ as

$$\boldsymbol{\theta}^t_{-i} = (\theta_1^{t+1}, \ldots, \theta_{i-1}^{t+1}, \theta_{i+1}^t, \ldots \theta_{21}^t).$$

For θ_i^{t+1} we sample a candidate value θ_i^* from the proposal distribution $q_i(\theta_i^* | \boldsymbol{\theta}^t_{-i}, \theta_i^t)$, where θ_i^t is the current value of the parameter θ_i. Then set $\theta_i^{t+1} = \theta_i^*$ with acceptance probability

$$\alpha(\boldsymbol{\theta}^t_{-i}, \theta_i^t, \theta_i^*) = \min\left(1, \frac{\pi(\theta_i^*) L(\boldsymbol{y} | \boldsymbol{\theta}^t_{-i}, \theta_i^*) q_i(\theta_i^t | \boldsymbol{\theta}^t_{-i}, \theta_i)}{\pi(\theta_i^t) L(\boldsymbol{y} | \boldsymbol{\theta}^t_{-i}, \theta_i^t) q_i(\theta_i | \boldsymbol{\theta}^t_{-i}, \theta_i^t)}\right), \tag{4.5.3}$$

where $\pi(\theta_i^*)$ and $\pi(\theta_i^t)$ are the prior distributions of the candidate and current values of the parameter θ_i, respectively, and $L(\boldsymbol{y} | \boldsymbol{\theta}^t_{-i}, \theta_i^*)$ and $L(\boldsymbol{y} | \boldsymbol{\theta}^t_{-i}, \theta_i^t)$ are the corresponding likelihood functions. The acceptance of the candidate value θ_i^* is determined by sampling a value, u, from the uniform distribution $U[0, 1]$. Thus, if $u < \alpha(\boldsymbol{\theta}^t_{-i}, \theta_i^t, \theta_i^*)$, then θ_i^{t+1} is set equal to θ_i^*; otherwise $\theta_i^{t+1} = \theta_i^t$. The iterations are continued for all parameters until the Markov chain reaches equilibrium at time T, where time before T is regarded as a burn-in period. Therefore, to obtain n draws from the posterior

distribution of the parameter θ_i, a further n updated simulations $\theta_i^{T+1}, \ldots, \theta_i^{T+n}$ are generated from the stationary distribution of the Markov chain. This sample is used to calculate different summary statistics for the posterior distribution of the parameter θ_i, such as the mean and 95% posterior intervals.

The convergence of the Markov chain can be investigated visually by plotting the values of the parameter against the number of iterations and inspecting the plot for signs of convergence. However, in our analysis we assess the convergence more formally by generating several independents chains, here three chains, and investigating the convergence by inspecting the mixing between the values of these chains visually by looking at the graph of the chains provided by WinBugs history in WinBugs software (Lunn and Spiegelhalter, 2000; version 1.4.3). We then compare the corresponding statistical summaries such as mean, median, and posterior intervals of the parameter obtained from each chain. A more formal test that uses the potential scale reduction factor $\widehat{R}$ of the parameter θ_i defined in equation 4.5.4 is used to assess the convergence of the chains to the required posterior distribution of the parameter.

$$\widehat{R}_i = \sqrt{\frac{\widehat{var}(\theta_i|\boldsymbol{y})}{W}}, \tag{4.5.4}$$

where

$$\widehat{var}(\theta_i|\boldsymbol{y}) = \frac{n-1}{n}W + \frac{1}{n}B,$$

where W and B are within and between chain variances of θ_i, and each chain is of length n after discarding the first half of the simulations, that is the burn-in period $T = n$ (Gelman et al., 2004, pp.296-297). This factor indicates the mixing index of the generated chains for the parameter θ_i, where a value of one indicates good mixing of the chains and convergence to the required posterior distribution of that parameter. A value of $\widehat{R}$ that is greater than one, particularly if $\widehat{R} > 1.1$, indicates that one would need to generate more simulations in order to reach equilibrium and improve the target distribution, since the scale of the sampling distribution of the underlying parameter decreases to one as $n \to \infty$. Having satisfied this condition for all parameters, that is

$\widehat{R} \rightarrow 1$ for all θ_i, one can combine the n simulations from each chain and regard them as a sample from the required posterior distribution of the underlying parameter.

Having specified the prior distributions and the MCMC method to sample from the posterior distribution, we use the R2WinBUGS package (Sturtz et al., 2005), that calls the WinBUGS software from the R package (R Core Team, 2013; version 3.0.2), to perform the Bayesian analysis for the TTO and DCE models. To sample from the posterior distribution of each parameter in the TTO and logit models under the suggested prior distributions: Gamma(1,10), Gamma(5,15), Uniform[0,1] and Beta(1,10), we generate three chains under each prior each with 10,000 iterations. To investigate the convergence of the chains to the required posterior distribution, we visually inspect the history of the generated chains using the sample monitoring tool in WinBUGS, and use the value of the mixing index of the chains to more formally investigate the convergence. If well-mixed, then we treat the second halves of these chains all together, i.e. 15,000 iterations, as a sample from the posterior distribution of the underlying parameters, and use this sample to calculate a summary of the posterior distribution for the parameter.

4.5.3 MCMC Results for TTO Model

For all the prior distributions, the scale reduction $\widehat{R}$ for each parameter in the TTO model appears to be near to 1, indicating convergence, as shown for the Gamma(1,10) prior in Figure 4.6.

Therefore, we regard the second halves of the three chains generated under each prior - 15,000 iterations from all chains - as a sample from the posterior distribution (see Appendix A.1 for the posterior distribution of each parameter). The posterior mean and 95% posterior intervals for each parameter in $\boldsymbol{\theta}$ are then estimated from their corresponding draws, as illustrated in Table 4.6

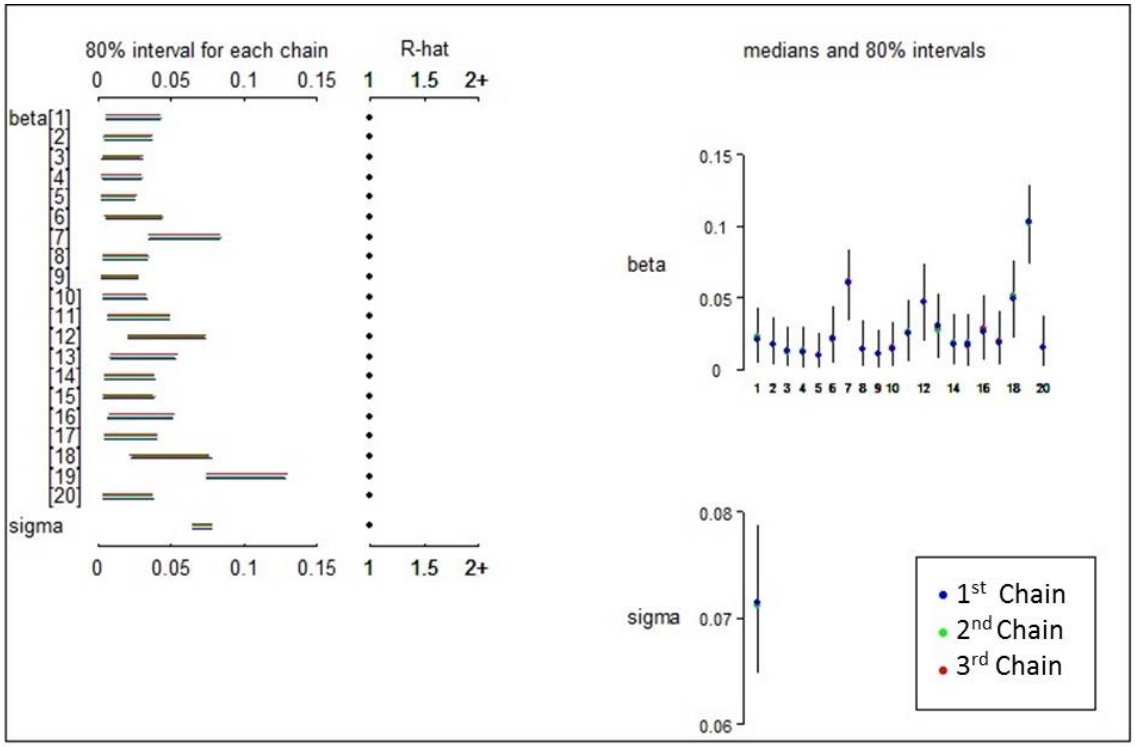

Figure 4.6: The mean and 80% posterior intervals for each parameter obtained from each chain, together with the mixing index of the chains $\widehat{R}$ using Gamma(1,10) prior

- **Model Parameters**

  For all proposed prior distributions, the OLS, i.e. the same as the m.l.e, differs slightly from the posterior mean. In general, a parameter with smaller m.l.e tends to have a larger posterior mean, and vice versa. Nevertheless, the m.l.e is still included within the range of the posterior mean, as shown by the 95% posterior intervals, except for the inconsistent estimates for level 1 and level 4 for the attributes breath and activities, respectively. Also, small m.l.e. values (e.g., β_3 and β_8) lie outside the posterior intervals of the parameters obtained using the Gamma(5,15). However, the general trend of the posterior mean for the parameters is consistent with the trend of the m.l.e., where the largest decrease in the mean utility is still associated with the change from level 2 to level 3 for the attribute activities.

  In terms of the sensitivity of the posterior inference to the choice of prior distribu-

Table 4.6: Mean and 95% posterior intervals for the TTO model parameters by prior distribution, together with the maximum likelihood estimators

Attribute	***Level***	***OLS***	***Gamma(1,10)***	***Gamma(5,15)***	***Uniform[0,1]***	***Beta(1,10)***
Concern	1	0.02989	0.0226(0.0013,0.0537)	0.0205(0.0073, 0.0383)	0.0224(0.0012, 0.0538)	0.0235(0.0013, 0.0557)
	2	0.02624	0.0193(0.0009, 0.0483)	0.0211(0.0078, 0.0391)	0.0195(0.0010, 0.0488)	0.0189(0.0009, 0.0476)
	3	0.00709	0.0146(0.0005, 0.0409)	0.0235(0.0087, 0.0434)	0.0146(0.0006, 0.0414)	0.0144(0.0006, 0.0406)
	4	0.00265	0.0144(0.0006, 0.0415)	0.0308(0.0116, 0.0562)	0.0156(0.0006, 0.0445)	0.0148(0.0005, 0.0425)
Short of Breath	1	-0.00845	0.0118(0.0004, 0.0360)	0.0184(0.0064, 0.0354)	0.0119(0.0004, 0.0369)	0.0121(0.0004, 0.0371)
	2	0.03000	0.0232(0.0012, 0.0575)	0.0262(0.0099, 0.0477)	0.0228(0.0012, 0.0567)	0.0229(0.0012, 0.0570)
	3	0.07615	0.0603(0.0204, 0.0972)	0.0439(0.0194, 0.0722)	0.0609(0.0214, 0.0982)	0.0601(0.0203, 0.0964)
	4	0.00090	0.0165(0.0006, 0.0484)	0.0341(0.0130, 0.0617)	0.0172(0.0006, 0.0499)	0.0166(0.0006, 0.0481)
Weather & pollution	1	0.00589	0.0127(0.0005, 0.0379)	0.0178(0.0062, 0.0340)	0.0126(0.0004, 0.0372)	0.0129(0.0005, 0.0378)
	2	0.01917	0.0169(0.0007, 0.0448)	0.0217(0.0079, 0.0401)	0.0168(0.0007, 0.0447)	0.0169(0.0007, 0.0440)
	3	0.02399	0.0268(0.0017, 0.0628)	0.0301(0.0117, 0.0540)	0.0269(0.0018, 0.0624)	0.0265(0.0018, 0.0608)
	4	0.05741	0.0475(0.0083, 0.0887)	0.0494(0.0214, 0.0823)	0.0494(0.0095, 0.0909)	0.0469(0.0083, 0.0880)
Sleep	1	0.04550	0.0301(0.0021, 0.0661)	0.0214(0.0079, 0.0396)	0.0291(0.0021, 0.0642)	0.0303(0.0022, 0.0662)
	2	0.01482	0.0200(0.0009, 0.0513)	0.0236(0.0088, 0.0430)	0.0204(0.0009, 0.0513)	0.0197(0.0009, 0.0498)
	3	0.01927	0.0192(0.0009, 0.0502)	0.0261(0.0098, 0.0476)	0.0192(0.0008, 0.0504)	0.0189(0.0009, 0.0499)
	4	0.02265	0.0285(0.0019, 0.0660)	0.0419(0.0172, 0.0723)	0.0300(0.0023, 0.0686)	0.0290(0.0021, 0.0670)
Activities	1	0.01244	0.0208(0.0009, 0.0556)	0.0239(0.0085, 0.0440)	0.0204(0.0009, 0.0535)	0.0200(0.0009, 0.0538)
	2	0.05134	0.0496(0.0096, 0.0912)	0.0454(0.0194, 0.0760)	0.0499(0.0097, 0.0922)	0.0501(0.0104, 0.0917)
	3	0.12366	0.1030(0.0584, 0.1444)	0.0774(0.0420, 0.1139)	0.1035(0.0585, 0.1466)	0.1023(0.0588, 0.1440)
	4	-0.00654	0.0181(0.0007, 0.0525)	0.0478(0.0193, 0.0826)	0.0194(0.0007, 0.0555)	0.0184(0.0007, 0.0524)
Scale	σ	0.07325	0.0717(0.0620, 0.0835)	0.0786(0.0670, 0.0929)	0.0721(0.0623, 0.0838)	0.0717(0.0617, 0.0838)

tion, Table 4.6 shows that posterior inferences, mean and 95% credible intervals, are similar for all parameters for the choice of uninformative prior uniform[0,1] and the more informative prior Gamma(1,10) and Beta(1,10) distributions. However, the posterior mean changes substantially when the prior mean moves away from zero as in the Gamma(5,15) prior, that is for a prior distribution with higher probability for larger values. The concentration of this prior at the larger values translates into a larger posterior mean for parameters with small values in the Gamma(1,10), Uniform[0,1] and Beta(1,10) priors. For illustration, in Figure 4.7 we present the posterior distribution for the parameters associated with levels 3 and 4 for the attribute activities (i.e. β_{19} and β_{20}) by the four prior distributions.

The plot shows that the posterior mean of the largest decrease in the mean utility, associated with a change from level 2 to level 3 for the attribute activities, β_{19}, decreases in Gamma(5,15) compared with those obtained when using Gamma(1,10), Uniform[0,1] and Beta(1,10) priors, whereas the posterior mean increases for the small decrease in the mean utility, β_{20}. Thus, if the researchers are sure about the range of the parameter values, but less confident whether the

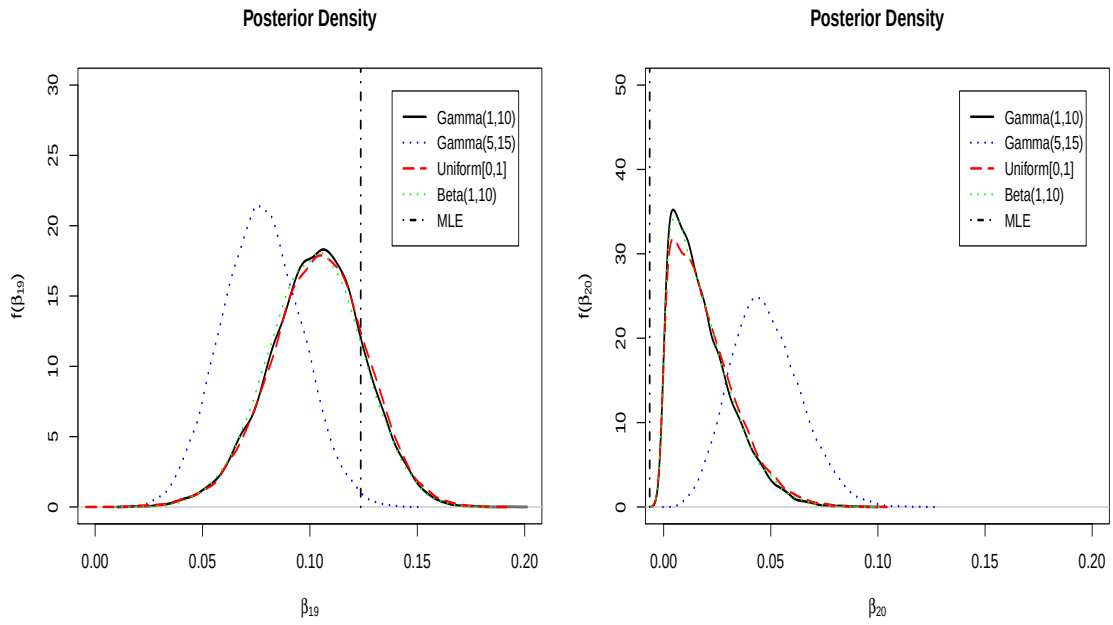

Figure 4.7: Posterior distribution for parameters associated with level 3 and 4 for the attribute activities, β_{19} and β_{20}, and the maximum likelihood estimates

parameter values lie very close to zero, it might be better to use the Uniform[0,1] prior. This is because the posterior distribution seems to be robust to the choice of prior distribution unless it has high probability for larger values.

The posterior means of the scale parameter, σ, are approximately similar under all prior distributions, and the m.l.e. of this parameter lies within the credible interval obtained under each prior. The posterior inferences of this parameter is less sensitive to the prior distribution chosen though the Gamma(5,15) still produces slightly more large value for this parameter as shown in Figure 4.8.

- **Health State Utilities**

  Using the 15,000 draws from the posterior distribution of each parameter, we calculate the posterior mean for each health state presented in the TTO exercise

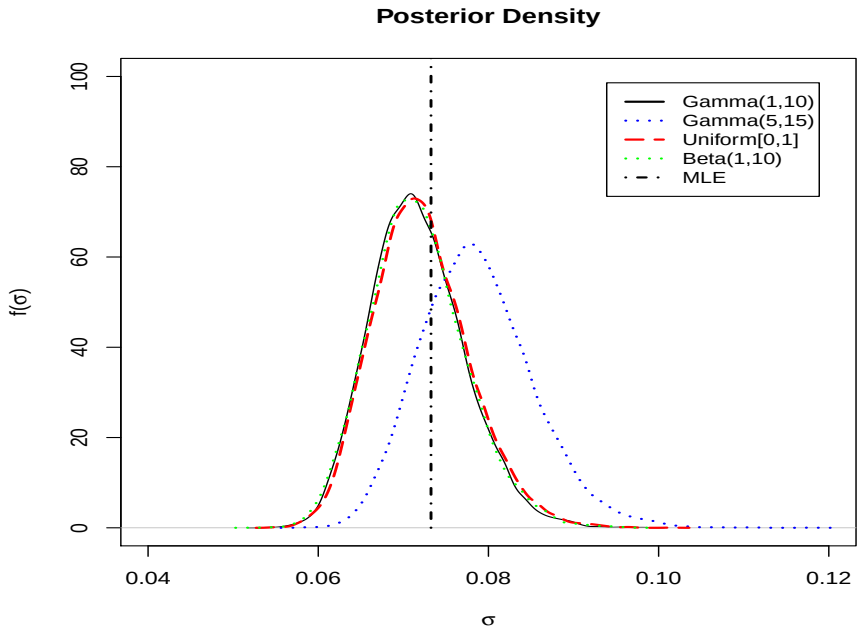

Figure 4.8: Posterior distribution of the scale parameter, σ, by the prior distribution together with the maximum likelihood estimate

as follows:

$$g(\boldsymbol{x}_{ij}) = \frac{1}{n}\sum_{n=1}^{15000}\left(1-\boldsymbol{\beta}_n\boldsymbol{x}_{ij}^T\right), \quad n=1,\ldots,15000, \quad j=1,\ldots,99, \tag{4.5.5}$$

where $\boldsymbol{\beta}_n = (\beta_{1,n},\ldots,\beta_{20,n})$ is the n draws from the posterior distribution of the parameters, and $\boldsymbol{x}_{ij}$ is a vector of dummy variables represents the health state as described in Section 4.3. Figure 4.9 shows the posterior mean utilities and 95% posterior intervals for each health state, together with the mean health state utilities computed using the m.l.e.

Comparing the maximum likelihood estimates with the posterior results, the plot illustrates that for Gamma(1,10), Uniform[0,1] and Beta(1,10) priors the posterior mean utilities for each health state are similar to those estimated using m.l.e, whereas, for Gamma(5,15), the posterior mean utilities are smaller than the maximum likelihood estimates. In general, the maximum likelihood estimates

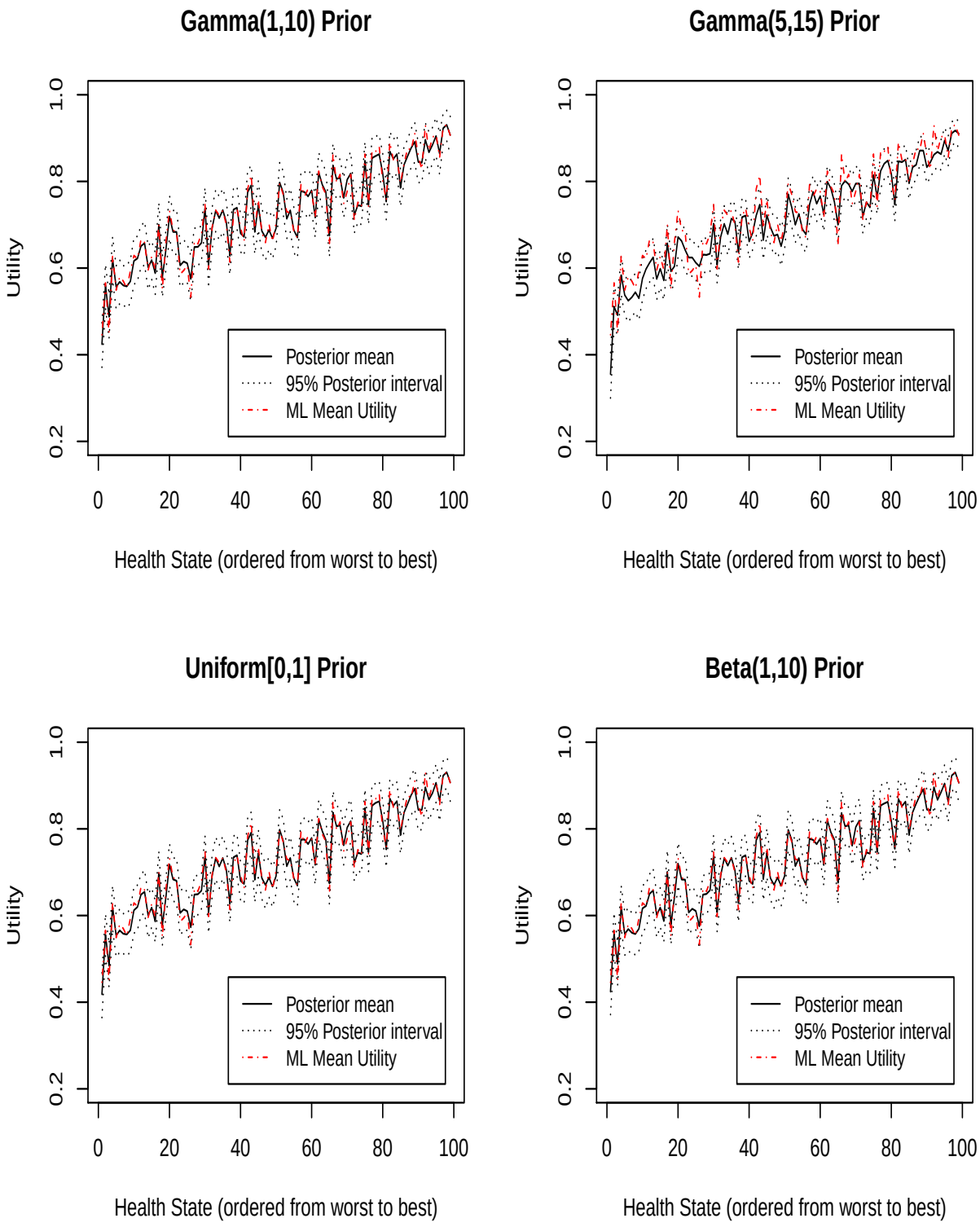

Figure 4.9: The mean and 95% posterior intervals for the mean health state utilities for the 99 health states used in the TTO survey by the prior distributions

for the utility values of those states are included in the 95% posterior intervals for Gamma(1,10), Uniform[0,1] and Beta(1,10) priors, whereas more of these estimates are excluded when Gamma(5,15) prior is used.

Comparing the results from different prior distributions, the posterior mean utilities for each health state are similar for Gamma(1,10), Uniform[0,1] and Beta(1,10) priors, but slightly different from those obtained using Gamma(5,15) prior, particularly for the severe health states. This is due to the fact that the Gamma(5,15) prior has large parameter estimates for the small decrease in mean utility, that is for parameters associated with change in the attributes from level 3 to level 4. This leads to larger decreases in the mean utility value, and, hence, produces smaller utility values for those states in comparison with those obtained when Gamma(1,10), Uniform[0,1] and Beta(1,10) priors are used.

For instance, consider the worst health state, where all the attributes are at the worst level (level 4 for each attribute). The posterior mean utility of this state is 0.355 when the Gamma(5,15) is used, whereas, for the Gamma(1,10) and the Uniform[0,1] priors the health state has a mean utility values of 0.4241, 0.4172, and 0.4243, respectively, and the posterior distributions are pulled further from zero, as illustrated in Figure 4.10.

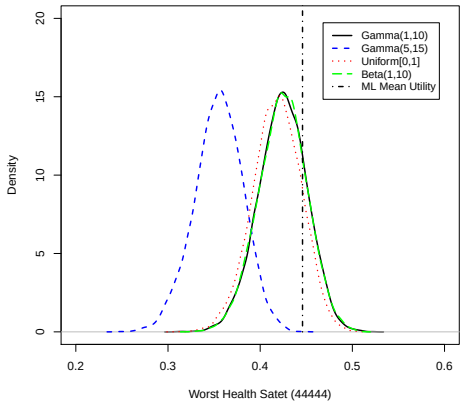

Figure 4.10: The posterior distributions for the worst health state defined by the AQL-5D classification system by the prior distributions

4.5.4 MCMC Results for the DCE Model

Similarly to the TTO data, for all prior distributions used in the analysis, the posterior distribution for each parameter in the logit model seems to converge to the true distribution as shown in Figure 4.11 by the mixing index of the chains $\widehat{R}$ for the Gamma(1,10) prior.

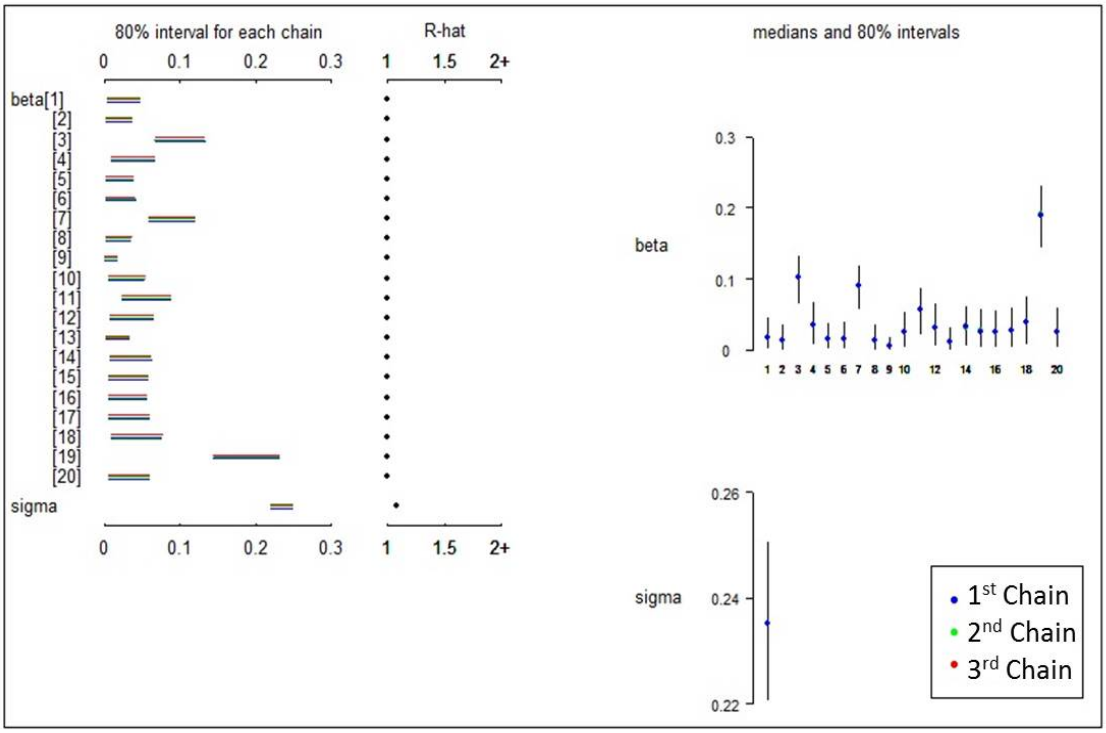

Figure 4.11: The mean and 80% posterior intervals for each parameter obtained from each chain, together with the mixing index of the chains $\widehat{R}$ using Gamma(1,10) prior

The plot shows that the values of the index appears to be close to one for all parameters, except for the scale parameter, σ, where $\widehat{R} = 1.07$. Nevertheless, this is still acceptable as it is less than 1.1, but it might require a higher level of precision in the final analysis (Gelman et al., 2004, p.297). Thus, we consider the second halves of the chains, i.e. 15,000 iterations in total, as a sample from the posterior distribution to compute different summary statistics for the parameters and estimate the posterior

Table 4.7: Mean and 95% posterior intervals for the logit model parameters by prior distributions, together with the maximum likelihood estimators

Attribute	*Level*	*m.l.e*	*Gamma(1,10)*	*Gamma(5,15)*	*Uniform[0,1]*	*Beta(1,10)*
Concern	1	0.00362	0.0223 (0.0008, 0.0661)	0.0369 (0.0133, 0.0686)	0.0223 (0.0008, 0.0656)	0.0228 (0.0008, 0.0675)
	2	$4.391e^{-07}$	0.0172 (0.0005, 0.0536)	0.0338 (0.0124, 0.0624)	0.0172 (0.0006, 0.0540)	0.0171 (0.0005, 0.0539)
	3	0.13354	0.1008 (0.0477, 0.1494)	0.0641 (0.0309, 0.1001)	00.1001 (0.0473, 0.1496)	0.1003 (0.0476, 0.1502)
	4	0.03375	0.0374 (0.0026, 0.0860)	0.0544 (0.0240, 0.0898)	0.0390 (0.0028, 0.0892)	0.0376 (0.0027, 0.0866)
Short of Breath	1	0.00421	0.0184 (0.0006, 0.0558)	0.0323 (0.0116, 0.0601)	0.0182 (0.0006, 0.0554)	0.0187 (0.0006, 0.0561)
	2	$2.645e^{-08}$	0.0194 (0.0006, 0.0590)	0.0337 (0.0126, 0.0627)	0.0193 (0.0006, 0.0595)	0.0193 (0.0006, 0.0591)
	3	0.12490	0.0902 (0.0414, 0.1350)	0.0603 (0.0289, 0.0950)	0.0903 (0.0416, 0.1361)	0.0899 (0.0419, 0.1340)
	4	0.00001	0.0168 (0.0005, 0.0513)	0.0371 (0.0140, 0.0678)	0.0177 (0.0006, 0.0543)	0.0171 (0.0006, 0.0530)
Weather & pollution	1	$7.070e^{-07}$	0.0080 (0.0002, 0.0282)	0.0215 (0.0073, 0.0426)	0.0080 (0.0002, 0.0278)	0.0081 (0.0002, 0.0283)
	2	0.02173	0.0284 (0.0014, 0.0707)	0.0344 (0.0133, 0.0617)	0.0275 (0.0013, 0.0697)	0.0284 (0.0014, 0.0711)
	3	0.07937	0.0563 (0.0089, 0.1058)	0.0450 (0.0191, 0.0770)	0.0577 (0.0096, 0.1090)	0.0557 (0.0092, 0.1052)
	4	0.02555	0.0339 (0.0017, 0.0864)	0.0445 (0.0173, 0.0788)	0.0356 (0.0018, 0.0882)	0.0342 (0.0018, 0.0862)
Sleep	1	$4.451e^{-06}$	0.0152 (0.0005, 0.0488)	0.0282 (0.0098, 0.0538)	0.0156 (0.0005, 0.0508)	0.0156 (0.0005, 0.0504)
	2	0.05593	0.0345 (0.0023, 0.0799)	0.0334 (0.0127, 0.0607)	0.0341 (0.0021, 0.0787)	0.0346 (0.0024, 0.0797)
	3	0.02225	0.0295 (0.0013, 0.0762)	0.0353 (0.0133, 0.0638)	0.0297 (0.0014, 0.0761)	0.0295 (0.0014, 0.0760)
	4	0.02927	0.0289 (0.0014, 0.0732)	0.0354 (0.0134, 0.0642)	0.0288 (0.0014, 0.0733)	0.0286 (0.0014, 0.0733)
Activities	1	0.03164	0.0314 (0.0015, 0.0792)	0.0427 (0.0163, 0.0766)	0.0311 (0.0015, 0.0801)	0.0317 (0.0016, 0.0796)
	2	0.02682	0.0415 (0.0025, 0.0973)	0.0618 (0.0264, 0.1033)	0.0417 (0.0025, 0.0998)	0.0417 (0.0026, 0.0985)
	3	0.24128	0.1887 (0.1195, 0.2526)	0.1189 (0.0682, 0.1707)	0.1898 (0.1204, 0.2553)	0.1874 (0.1187, 0.2520)
	4	0.00017	0.0300 (0.0013, 0.0817)	0.0537 (0.0216, 0.0930)	0.0306 (0.0013, 0.0823)	0.0307 (0.0013, 0.0825)
Scale	σ	0.24140	0.2392(0.2138, 0.2597)	0.2254 (0.2057, 0.2477)	0.2363 (0.2143, 0.2610)	0.2351 (0.2134, 0.2595)

mean utilities.

- **Model Parameters**

  Table 4.7 shows the mean and 95% posterior intervals for each parameter in the logit model. Similar to the TTO model, the posterior mean has a comparatively large value when the maximum likelihood estimates are small and vice versa. Also, most of the maximum likelihood estimates are included within the 95% posterior intervals in Gamma(1,10), Uniform[0,1] and Beta(1,10) priors, but several more fall outside the posterior intervals when the Gamma(5,15) prior is used; more so than for the TTO model, that is 12 parameters compared with 8.

  Once again as in the TTO model, the posterior distributions for each parameter are similar when using Gamma(1,10) Uniform[0,1] and Beta(1,10) priors, but change dramatically for the Gamma(5,15) priors, as shown for the parameter associated with the largest decrease in the mean utility, i.e. change from level two to level three in the activities attribute, β_{19}, in the left panel of Figure 4.12. Similarly for the parameters with large values for the m.l.e such as the parameters associated with level three for attributes concern and short of breath, i.e. β_3 and β_7, respectively. The figure also presents the posterior distributions

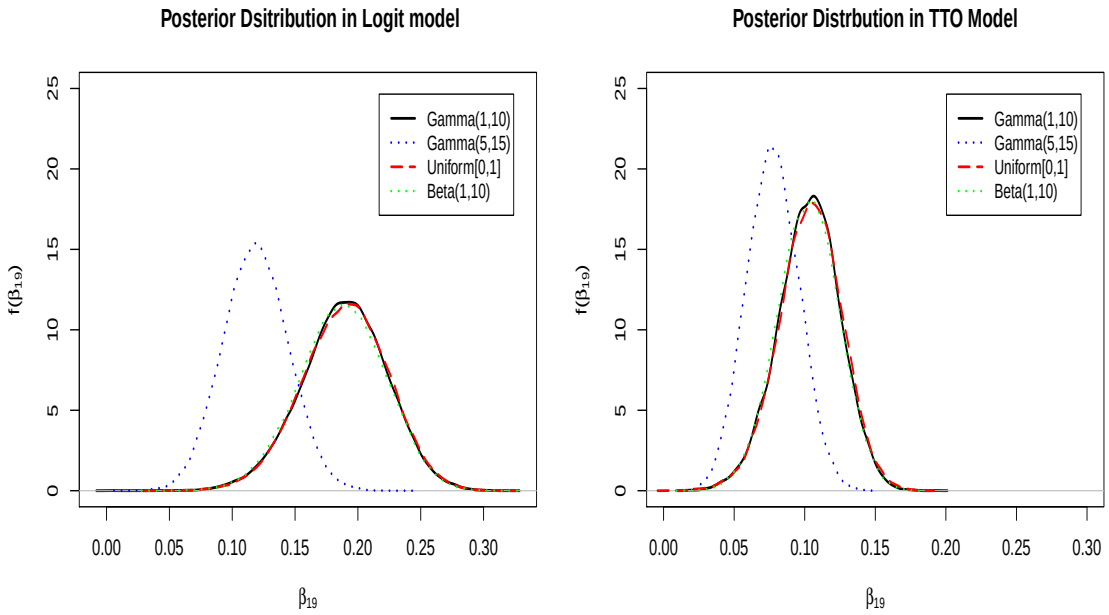

Figure 4.12: The posterior distributions for β_{19} using logit and TTO models by the prior distributions

for this parameter using the TTO models. The plot shows that the change in the posterior distribution with respect to Gamma(5,15) prior, i.e. the prior with high probability for the large values, is larger in the logit model than in the TTO model. Thus, the logit model might be more sensitive to the prior distribution with higher probability for large value than the TTO model.

Comparing the mean and posterior intervals for each parameter using the logit and TTO Bayesian models, we observe that, in general, the posterior mean in the TTO model is lower than those obtained in the logit model, as shown in Figure 4.13 for the Gamma(1,10) prior. However, as discussed earlier, the logit model is not expected to produce the same results as the TTO model. Therefore, we compare the parameter uncertainty produced by each model instead using the 95% credible interval of the posterior mean of the parameters, where the wider the interval the more uncertainty in the posterior mean values. Considering the width of the posterior intervals of each parameter under the Gamma(1,10) prior shown

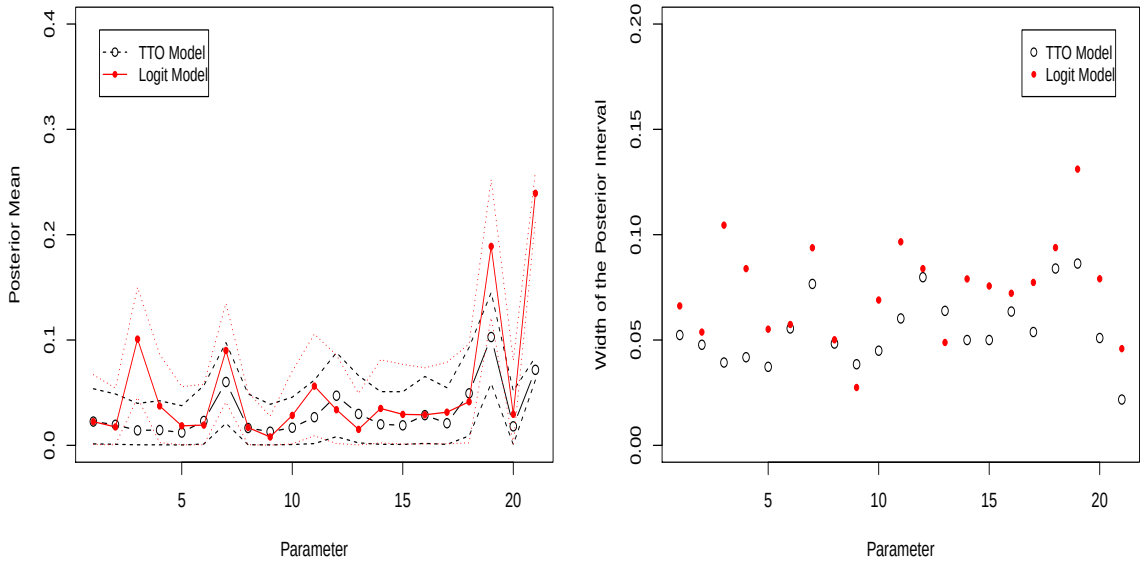

Figure 4.13: Mean and 95% posterior intervals together with the widths of the intervals for each parameter using the logit and TTO models and Gamma(1,10) prior

in the right panel of Figure 4.13, it can be seen that the logit model produces a slightly larger uncertainty in the posterior mean values compared to the TTO model, particularly for those parameter associated with larger change in the mean utility values such as β_3, β_4 and β_{19} as well as the scale parameter σ, where the logit model results in a wider credible interval. Similar results are obtained under the other prior distributions, i.e. the Gamma(5,15) and uniform[0,1] priors. Nevertheless, this might be related to the fact that the TTO survey has more respondents and observations compared to the DCE exercise, as shown in Table 4.5. This would result in smaller variation in the collected data and, hence, more precise estimates for the preference parameters and the mean utilities.

- **Health State Utilities**

  The mean utilities for each health state presented in the pairwise choice experiment survey are calculated using the 15,000 iterations from the posterior

distribution of the parameters as follows:

$$g(\boldsymbol{x}_{ijs}) = \frac{1}{n}\sum_{n=1}^{15000}\left(1-\boldsymbol{\beta}_n\boldsymbol{x}_{ijs}^T\right), \quad n=1,\ldots,15000, \quad j=1,2, \quad s=1,\ldots,32, \tag{4.5.6}$$

where $\boldsymbol{\beta}_n$ is the n^{th} draw from the posterior distribution of the parameter vector $\boldsymbol{\beta}$. Figure 4.14 presents the mean and 95% posterior intervals for these health states (excluding death state which has a utility of zero) using different prior distributions, together with the estimated utilities values using the maximum likelihood estimates. The plot shows that, generally for all prior distributions, the posterior mean utilities are lower than utility values estimated using the m.l.e. Nevertheless, the mean health state utilities estimated using the maximum likelihood estimates are included within the 95% posterior intervals when Gamma(1,10), Uniform[0,1] and Beta(1,10) priors are used, whereas most of these estimates fall outside the 95% posterior intervals in Gamma(5,15) prior. In addition, as in the TTO model, the plot shows that the posterior distribution for the mean utilities are similar when using Gamma(1,10), Uniform[0,1] and Beta(1,10) priors, but is dramatically different when Gamma(5,15) prior is used.

To compare the predictive ability of the TTO and logit models, we first predict the value of the 99 health states presented in the TTO survey using the logit model and then compare it with those obtained using the TTO model. This is because it is not possible to directly compare states in the TTO and DCE surveys as they evaluate different health states. In our analysis, we expect that the logit model would produce lower posterior mean utility values compared to the TTO model particularly for the severe states. Comparing the predicted posterior mean utilities of the 99 states shown in Figure 4.15 for the TTO and logit models, it can be seen that the mean utility values are pulled more toward zero in the logit model compared to the TTO model particularly for the more sever states (e.g. the worst health state 44444). This follows since the logit model produces larger values for the posterior mean of the parameters particularly those associated with

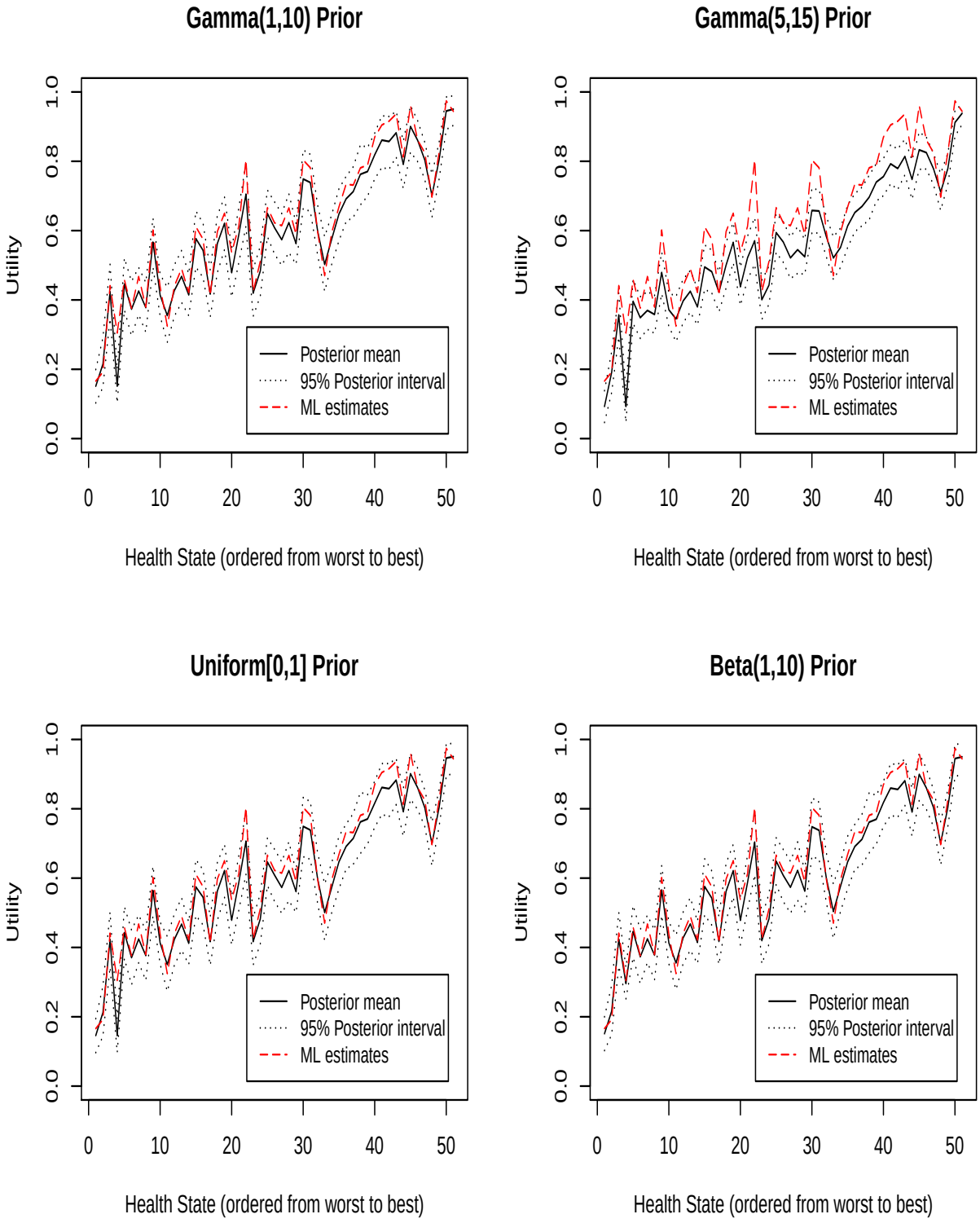

Figure 4.14: Mean and 95% posterior intervals for the utility values for the health states evaluated in the DCE survey, excluding the death state, by different prior distributions

the change from level 2 to level 3 in each attribute (e.g., β_3 and β_{19}) as show in Figures 4.13.

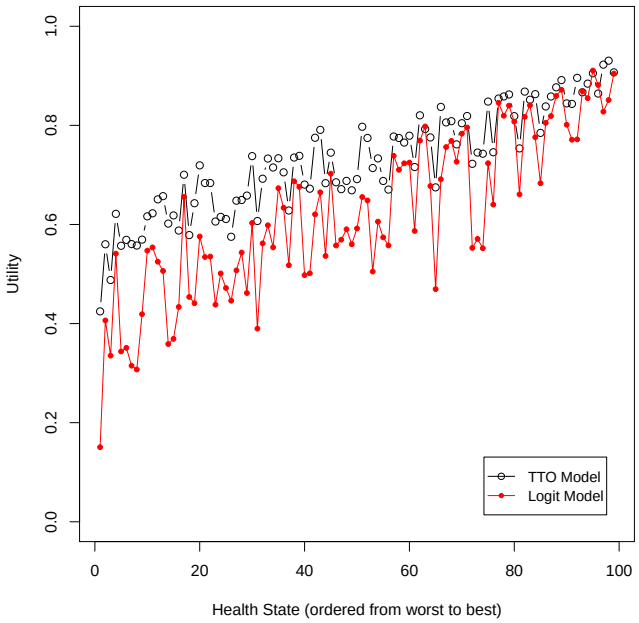

Figure 4.15: The posterior mean utilities for 99 health states presented in the TTO survey estimated using logit and TTO models by the Gamma(1,10) prior

Now, since either TTO or logit model can be used to estimate the mean utility of any health state defined by the classification system, both models are not expected to produce the same results. Therefore, we compare the uncertainty in the mean utility values produce by both models instead of comparing the absolute utility values. This can be done by comparing the 95% posterior intervals of the posterior mean utilities in both models, which can be represented more clearly using the width of those intervals as shown in Figure 4.16 for the Gamma(1,10) prior distribution.

The plot shows that the DCE data seems to produce larger uncertainty, particularly for the more severe health states, as it results in wider posterior interval that implies larger uncertainty. Nevertheless, for the worst health state (AQL-5D

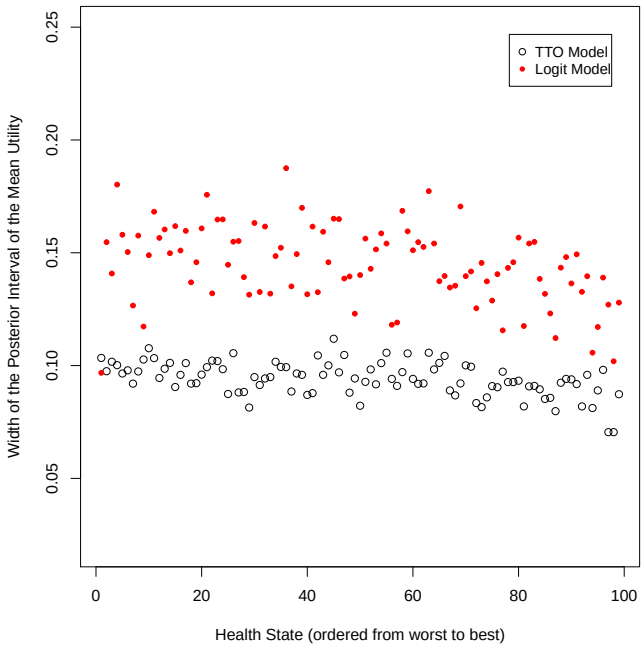

Figure 4.16: The width of the posterior interval of the mean utilities of the 99 health states presented in the TTO survey using logit and TTO models by the Gamma(1,10) prior

health state 44444) the logit model seems to have slightly less uncertainty around the posterior mean utility value compared to the TTO model, similarly for the other priors. This might be related to the fact that the TTO values, as opposed to the DCE methods, are affect by time preference and require altering the choice task when the state is consider worse than death. This might affect the average mean utility value of this state and increase the variations in the respondents' valuations of this states, where some respondents might consider this state as worth living and not willing to trade much of their life expectancy for being in a healthier state while other might consider it worse than being dead.

4.6 Summary and Discussion

In this chapter, we analysed two data sets, TTO and pairwise comparisons data, using classical and Bayesian approaches. The ordinary linear model and logit model were fitted to the TTO and DCE data, respectively. In general, the logit model produces higher maximum likelihood estimates for those parameters associated with more severe levels of the attributes (i.e. level 3 and 4), and hence results in lower estimates for the mean utility values for the most severe health states defined by the AQL-5D classification system.

The Bayesian approach has been used to obtain probability distributions for the preference parameters and the health state utility values, as they are usually required by health economists to perform a probabilistic sensitivity analysis to account for the uncertainty in their decision. Also, in this chapter, we analysed the data using the Bayesian approach to obtain posterior distributions for the unknown logit model parameters, and then use this information about the parameters when constructing the choice design for the same classification system. In the Bayesian analysis, we used the MCMC method to sample from the posterior distribution of each parameter in the model, since we cannot derive the posterior distribution analytically. Different prior distributions were considered to investigate the robustness of the Bayesian inference to the choice of prior. In particular, Gamma(1,10), Gamma(5,15), Uniform[0,1] and Beta(1,10) priors were used to investigated the effect of using informative and less informative priors, such as the Uniform[0,1] where all the parameters have an equal probability of being anywhere between 0 and 1. Also, the Gamma(5,15) prior has been used to investigate the effect of the prior distribution with higher probability for the larger values in the posterior inference.

For both the TTO and logit models, the posterior mean utilities were smaller than those obtained using the maximum likelihood estimates as shown in Figures 4.9 and 4.14, particularly when a Gamma(5,15) prior is used. The Gamma(1,10), Uniform[0,1] and Beta(1,10) priors produce similar mean utilities and posterior distributions, whereas the results change dramatically when the Gamma(5,15) was used, particularly in the logit model. Thus, the posterior inference is robust to the choice of prior unless the prior used has a high probability for larger parameter values. However, since in reality some parameters are expected to produce a larger decrease in the mean utility than others, specifying different priors for each parameter might be more appropriate. That is, for small parameter values, it would be more appropriate to use a prior distribution that favours small values and vice versa.

Comparing the uncertainty produced by both TTO and DCE models, generally, the uncertainty in the parameter and utility values is larger in the logit model compared to the TTO model. Nevertheless, this might be related to the number of respondents and choice data used in each survey, since the TTO survey has more respondents and observations than the DCE. This would result in more precise estimates for the preference parameters, and consequently the mean utilities. Though the TTO data might provide less uncertainty in the parameters and the utility values here; they might not reflect the true values for these quantities. This is because TTO data are affected by other non-health factors such as time preference, and also it might be harder for respondents to perform than the pairwise choices.

Therefore, in practice, DCEs might be more appropriate to value health state utility as they are not effected by non-health factors and they are simpler than the TTO valuation technique. Nevertheless, we cannot rule out the impact of the choice of the DCE design on the results of the DCE data, since obtaining reliable parameter values depends on the information collected from the DCE survey, which itself depends in the way that the choice design is constructed. Here, the DCE survey used a limited number of choice sets that are constructed based on the four principles of the optimal

choice design developed in Huber and Zwerina (1996): level balance, orthogonality, minimal overlap, and utility balance (that is satisfied in level balanced design used here by assuming zero point priors for the preference parameters). However, this design method, as discussed in Section 3.2.3, does not guarantee obtaining an optimal design, or the ability to estimate the main effects of interest. Also, the assumption of zero priors produces some dominant choices in the DCE survey (three dominant choices). This will reduce the efficiency of the choice design, and have an impact on the results and might increase the uncertainty in the parameter and the utility values.

A more sophisticated choice design approach might improve the quality of the collected data from the DCE, and hence the logit model results. In the design literature, there has been concern regarding experimental design methods used for DCEs and some developments to produce more efficient choice designs, as shown for the Bayesian design literature in Chapter 3. Therefore, in the following chapter, we concentrate on improving the DCE design for valuing health state utilities within the QALY framework, particularly for the AQL-5D case study, using these developments such that it produces more reliable values for the parameter estimates and hence the utility values. In particular, using a Bayesian optimal choice design approach, that accounts for uncertainty in the unknown model parameters through incorporating a prior information about the parameters in constructing the choice design.

Chapter 5

Bayesian Optimal Choice Designs for Valuing Health State Utilities

5.1 Introduction

In this chapter, we present our design algorithm to construct Bayesian optimal designs for valuing health states within the QALY framework. The developed choice design algorithm is then used to construct Bayesian optimal pairwise comparison designs for the AQL-5D classification system. We compare the resulting Bayesian designs to the level balanced design (LBD) introduced in Chapter 4.

Constructing Bayesian optimal designs, as discussed in Chapter 3, requires specifying the choice model, the design criterion and a prior distribution for the unknown model parameters. Therefore, we next derive the Bayesian $\mathcal{D}_{\mathbf{S}}$-optimality criterion for our design model, i.e. the multinomial logit (MNL) model. This criterion is chosen

such that the volume of the posterior credible ellipsoid of the unknown preference parameters, $\boldsymbol{\beta}$, is minimised, and hence more reliable parameter estimates and health state utility values are obtained. In Section 5.3, we discuss the problems with the available software to construct Bayesian choice designs for valuing health states. We introduce our design algorithm in Section 5.4, and demonstrate how the choice task could be simplified to reduce the response errors in Section 5.5. The developed choice algorithm together with the simplification of the choice task are then applied to generate Bayesian optimal pairwise comparisons for the AQL-5D, in which the resulting designs are compared with each other and with the LBD as illustrated in Section 5.6. A summary and discussion of the main findings in this chapter is presented in Section 5.7.

5.2

Deriving the Optimality Criterion for the MNL model

In this section, we derive the appropriate form of the Bayesian $\mathcal{D}$-optimality criterion for the MNL model including the death state, using the asymptotic normal approximation to the posterior distribution for the unknown model parameters as described in Section 3.6.1. We base our derivation of the design criterion on the work of Kessels et al. (2004) who derived the Bayesian $\mathcal{D}$-optimality criterion for the general MNL model – the MNL model without the death state – using the asymptotic approximation for the posterior distribution.

The posterior variance-covariance matrix (VCM) of this asymptotic distribution is

approximated by the inverse of the Fisher information matrix (FIM) or the generalised Fisher information matrix (GFIM), where the Bayesian $\mathcal{D}$-criterion seeks to minimise the determinant of the VCM to minimise the posterior credible ellipsoid of the model parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\sigma})$. However, the goal of performing a choice experiment, here, is to estimate the preference parameters, $\boldsymbol{\beta}$, as precisely as possible, therefore our interest is in a subset of the MNL model parameters. This interest should be reflected in the $\mathcal{D}$-optimal design criterion as discussed in Goos et al. (2010) who use the $\mathcal{D}_{\mathbf{S}}$-optimality criterion to construct their choice design. Therefore, an appropriate design criterion will be the $\mathcal{D}_{\mathbf{S}}$-optimality criterion which seeks designs that minimise the variance of the parameters of interest, here the preference parameters, $\boldsymbol{\beta}$. The derivation of this optimality design criterion requires partitioning the information matrix as

$$\begin{aligned} \text{FIM}(\xi, \boldsymbol{\theta}) &= \begin{bmatrix} \text{FIM}^{11}(\xi, \boldsymbol{\theta}) & \text{FIM}^{12}(\xi, \boldsymbol{\theta}) \\ \text{FIM}^{21}(\xi, \boldsymbol{\theta}) & \text{FIM}^{22}(\xi, \boldsymbol{\theta}) \end{bmatrix}, \\ &= \begin{bmatrix} -E_{\mathbf{Y}}\left\{\frac{\partial^2 l(\mathbf{y}|\boldsymbol{\beta},\sigma)}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}}\right\} & -E_{\mathbf{Y}}\left\{\frac{\partial^2 l(\mathbf{y}|\boldsymbol{\beta},\sigma)}{\partial \sigma \partial \boldsymbol{\beta}}\right\} \\ -E_{\mathbf{Y}}\left\{\frac{\partial^2 l(\mathbf{y}|\boldsymbol{\beta},\sigma)}{\partial \boldsymbol{\beta}^T \partial \sigma}\right\} & -E_{\mathbf{Y}}\left\{\frac{\partial^2 l(\mathbf{y}|\boldsymbol{\beta},\sigma)}{\partial \sigma^2}\right\} \end{bmatrix}, \end{aligned} \tag{5.2.1}$$

where $l(\mathbf{y}|\boldsymbol{\beta}, \sigma)$ is the log likelihood function of the MNL defined as in Equation (2.5.17):

$$l(\mathbf{y}|\boldsymbol{\beta}, \sigma) = \sum_{s=1}^{S} \sum_{j=1}^{J} \sum_{i=1}^{N} y_{ijs} \log P_{ijs},$$

where y_{ijs} is a dummy variable that equals 1 if health state $\boldsymbol{x}_{ijs}$ is chosen and is 0 otherwise.

The posterior VCM of the parameters $\boldsymbol{\beta}$ can be estimated asymptotically using the results on the inverse of a partitioned matrix as a submatrix of the inverse of $\text{FIM}(\xi, \boldsymbol{\theta})$ (Atkinson et al., 2007) as

$$\widehat{Var}(\boldsymbol{\beta}) = \left\{\text{FIM}^{11}(\xi, \boldsymbol{\theta}) - \text{FIM}^{12}(\xi, \boldsymbol{\theta}) \left[\text{FIM}^{22}(\xi, \boldsymbol{\theta})\right]^{-1} \text{FIM}^{21}(\xi, \boldsymbol{\theta})\right\}^{-1}. \tag{5.2.2}$$

Hence, the $\mathcal{D}_{\mathbf{S}}$-optimality design criterion is defined as the determinant of this inverse, where $\mathcal{D}_{\mathbf{S}}$-optimal design is obtained by minimising the determinant

$$\left|\left\{\mathrm{FIM}^{11}(\xi,\boldsymbol{\theta}) - \mathrm{FIM}^{12}(\xi,\boldsymbol{\theta})\left[\mathrm{FIM}^{22}(\xi,\boldsymbol{\theta})\right]^{-1}\mathrm{FIM}^{21}(\xi,\boldsymbol{\theta})\right\}^{-1}\right|, \tag{5.2.3}$$

or, equivalently maximising the determinant of the inverse of the posterior VCM of the preference parameters, that is the amount of information a design ξ carries about the unknown preference parameters in the MNL model, $\boldsymbol{\beta}$, denoted as $\mathrm{FIM}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta})$,

$$\mathrm{FIM}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta}) = \mathrm{FIM}^{11}(\xi,\boldsymbol{\theta}) - \mathrm{FIM}^{12}(\xi,\boldsymbol{\theta})\left[\mathrm{FIM}^{22}(\xi,\boldsymbol{\theta})\right]^{-1}\mathrm{FIM}^{21}(\xi,\boldsymbol{\theta}). \tag{5.2.4}$$

Thus, to derive the Bayesian $\mathcal{D}_{\mathbf{S}}$-optimality design criterion for the MNL model, we need to calculate the components of the partitioned FIM or the corresponding GFIM for the MNL model, and then deduce the optimality criterion for the binary logit model to construct a pairwise experiment.

In a choice experiment each individual i is presented with a set of health states to choose from, $C_s = \{\boldsymbol{x}_{i1s}, \ldots, \boldsymbol{x}_{iJs}\}$, where each health state is represented by a vector of dummy variables $\boldsymbol{x}_{ijs}$ with elements defined as

$$x_{\lambda\delta} = \begin{cases} 1 & \text{if attribute } \delta \text{ of health state } \boldsymbol{x}_{ij} \text{ is at level } \lambda \text{ or higher,} \\ 0 & \text{otherwise.} \end{cases}$$

For valuing health states within the QALY scale, an individual may also be asked to evaluate a set of health states that includes the death state $C_s = \{\boldsymbol{x}_{i1s}, \ldots, \boldsymbol{x}_{i(J-1)s}, \boldsymbol{x}_d\}$, in which case $\boldsymbol{x}_{iJs} = \boldsymbol{x}_d$, where $\boldsymbol{x}_d$ is a vector of a dummy variable with elements of 0 corresponding to each attribute level differences and a last element of 1 representing the death state as defined in Section 2.3.3. In Section 2.5.3, we defined the MNL model including the death state, where the MNL choice probability that individual i chooses

health state $\boldsymbol{x}_{ijs}$ in a non-death choice set, i.e. $\boldsymbol{x}_d \notin C_s$, is derived as

$$P_{ijs} = \frac{\exp\left(\frac{g(\boldsymbol{x}_{ijs})}{\sigma}\right)}{\sum_{t=1}^{J} \exp\left(\frac{g(\boldsymbol{x}_{its})}{\sigma}\right)},$$

where $g(\boldsymbol{x}_{ijs}) = 1 - \boldsymbol{\beta}\boldsymbol{x}_{ijs}^T$ is the population mean utility for health state $\boldsymbol{x}_{ijs}$, and $\boldsymbol{\beta}$ is a vector of the unknown model parameters associated with the incremental decrease in the mean utility when moving one level within one attribute, as defined in Equation (2.3.1). The parameter σ is the scale parameter of the random component of the utility, ϵ_{ijs}.

For health state $\boldsymbol{x}_{ijs}$ in a choice set that includes death, i.e. $\boldsymbol{x}_d \in C_s$, the MNL choice probability is derived as

$$P_{ijs} = \frac{\exp\left(\frac{g(\boldsymbol{x}_{ijs})}{\sigma}\right)}{\sum_{t=1}^{J-1} \exp\left(\frac{g(\boldsymbol{x}_{its})}{\sigma}\right)} \left\{1 - \exp\left[-\sum_{t=1}^{J-1} \exp\left(\frac{g(\boldsymbol{x}_{its}) + \mu}{\sigma}\right)\right]\right\},$$

and for the death state it is given by

$$P_{iJs} = \exp\left[-\sum_{t=1}^{J-1} \exp\left(\frac{g(\boldsymbol{x}_{its}) + \mu}{\sigma}\right)\right].$$

The components of the FIM for a design ξ are computed by deriving the sub-Fisher information matrices (sub-FIMs) defined in Equation (5.2.1), that is computed as the negative value of the expected second derivative of the log likelihood function defined in Equation 2.5.17. Since the MNL model assumes homogenous preference across respondents (i.e. individuals have the same choice probability for the alternative $\boldsymbol{x}_{ijs}$), we can drop the i index from the choice probability. Also, to simplify the calculation, we start by computing the sub-FIMs defined in Equation (5.2.1) for one respondent (i.e. taking $N = 1$), and then multiply it by the total number of respondents N, as in Kessels et al. (2004). Thus, the sub-FIMs for one respondent, which we denote by

FIM_I, can be written as

$$\text{FIM}_I^{11}(\xi, \boldsymbol{\theta}) = -E_{\mathbf{Y}} \left\{ \sum_{s=1}^{S} \sum_{j=1}^{J} y_{js} \frac{\partial^2}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} \log(P_{js}) \right\}, \tag{5.2.5}$$

$$\text{FIM}_I^{12}(\xi, \boldsymbol{\theta}) = -E_{\mathbf{Y}} \left\{ \sum_{s=1}^{S} \sum_{j=1}^{J} y_{js} \frac{\partial^2}{\partial \sigma \partial \boldsymbol{\beta}} \log(P_{js}) \right\}, \tag{5.2.6}$$

$$\text{FIM}_I^{22}(\xi, \boldsymbol{\theta}) = -E_{\mathbf{Y}} \left\{ \sum_{s=1}^{S} \sum_{j=1}^{J} y_{js} \frac{\partial^2}{\partial \sigma^2} \log(P_{js}) \right\}, \tag{5.2.7}$$

and $\text{FIM}_I^{21}(\xi, \boldsymbol{\theta}) = \left[\text{FIM}_I^{12}(\xi, \boldsymbol{\theta})\right]^T$.

To evaluate these matrices, we need to compute the derivatives for the corresponding choice probabilities in each choice set. The derivative is taken with respect to the parameter vector $\boldsymbol{\beta}$ and σ after substituting the definition of the population mean utility, $g(\boldsymbol{x}_{ijs}) = 1 - \boldsymbol{\beta}\boldsymbol{x}_{ijs}^T$, in the defined choice probabilities. Therefore, we start by computing the derivative for the MNL choice probabilities in non-death choice sets with respect to $\boldsymbol{\beta}$ and σ as

$$\begin{aligned}
\frac{\partial^2}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} \log(P_{js}) &= \frac{\partial}{\partial \boldsymbol{\beta}^T} \left\{ \frac{\partial}{\partial \boldsymbol{\beta}} \log \left[\frac{\exp\left(\frac{g(\boldsymbol{x}_{js})}{\sigma}\right)}{\sum_{t=1}^{J} \exp\left(\frac{g(\boldsymbol{x}_{ts})}{\sigma}\right)} \right] \right\}, \\
&= \frac{\partial}{\partial \boldsymbol{\beta}^T} \left\{ \frac{1}{\sigma} \left[-\boldsymbol{x}_{js}^T + \frac{\sum_{t=1}^{J} \exp\left(\frac{g(\boldsymbol{x}_{ts})}{\sigma}\right) \boldsymbol{x}_{ts}^T}{\sum_{t=1}^{J} \exp\left(\frac{g(\boldsymbol{x}_{ts})}{\sigma}\right)} \right] \right\}, \\
&= -\frac{1}{\sigma^2} \left(\sum_{t=1}^{J} P_{ts} \boldsymbol{x}_{ts}^T \boldsymbol{x}_{ts} - \sum_{t=1}^{J} P_{ts} \boldsymbol{x}_{ts}^T \sum_{t=1}^{J} P_{ts} \boldsymbol{x}_{ts} \right), \\
&= -\frac{1}{\sigma^2} \left[\mathbf{X}_s (\mathbf{P}_s - \mathbf{p}_s \mathbf{p}_s^T) \mathbf{X}_s^T \right],
\end{aligned} \tag{5.2.8}$$

where $\mathbf{X}_s = [\boldsymbol{x}_{1s}^T, \ldots, \boldsymbol{x}_{Js}^T]$ is the design matrix for choice set s. The terms $\mathbf{P}_s$ and $\mathbf{p}_s$ are a $J \times J$ diagonal matrix and a $J \times 1$ vector, respectively. Element j, j of $\mathbf{P}_s$ and element j of $\mathbf{p}_s$ are both equal to $P_{js} = \frac{\exp\left(\frac{g(\boldsymbol{x}_{js})}{\sigma}\right)}{\sum_{t=1}^{J} \exp\left(\frac{g(\boldsymbol{x}_{ts})}{\sigma}\right)}$.

Taking the derivative with respect to $\boldsymbol{\beta}$ and then σ gives

$$\begin{aligned}
\frac{\partial^2}{\partial\sigma\partial\boldsymbol{\beta}} \log(P_{js}) &= \frac{\partial}{\partial\sigma} \left\{ \frac{\partial}{\partial\boldsymbol{\beta}} \log \left[\frac{\exp\left(\frac{g(\boldsymbol{x}_{js})}{\sigma}\right)}{\sum_{t=1}^{J} \exp\left(\frac{g(\boldsymbol{x}_{ts})}{\sigma}\right)} \right] \right\}, \\
&= \frac{\partial}{\partial\sigma} \left\{ \frac{1}{\sigma} \left[-\boldsymbol{x}_{js}^T + \sum_{t=1}^{J} P_{ts} \boldsymbol{x}_{ts}^T \right] \right\}, \\
&= \frac{1}{\sigma^3} \left[\sigma(\boldsymbol{x}_{js}^T - \mathbf{X}_s \mathbf{p}_s) - \mathbf{X}_s(\mathbf{P}_s - \mathbf{p}_s \mathbf{p}_s^T)\mathbf{U}_s^T \right], \qquad (5.2.9)
\end{aligned}$$

where $\mathbf{X}_s$, $\mathbf{P}_s$ and $\mathbf{p}_s$ defined as in Equation (5.2.8), and $\mathbf{U}_s = [g(\boldsymbol{x}_{1s}), \ldots, g(\boldsymbol{x}_{Js})]$ is a row vector of the mean utility value of each alternative in choice set s.

The second derivative of the choice probability with respect to the scale parameter σ is derived as

$$\begin{aligned}
\frac{\partial^2}{\partial\sigma^2} \log(P_{js}) &= \frac{\partial}{\partial\sigma} \left\{ \frac{\partial}{\partial\sigma} \log \left[\frac{\exp\left(\frac{g(\boldsymbol{x}_{js})}{\sigma}\right)}{\sum_{t=1}^{J} \exp\left(\frac{g(\boldsymbol{x}_{ts})}{\sigma}\right)} \right] \right\}, \\
&= \frac{\partial}{\partial\sigma} \left\{ \frac{1}{\sigma^2} \left(-g(\boldsymbol{x}_{js}) + \sum_{t=1}^{J} P_{ts} g(\boldsymbol{x}_{ts}) \right) \right\}, \\
&= \frac{1}{\sigma^4} \left[2\sigma g(\boldsymbol{x}_{js}) - \sum_{t=1}^{J} P_{ts} g(\boldsymbol{x}_{ts}) \left(2\sigma + g(\boldsymbol{x}_{ts}) - \right.\right. \\
&\quad \left.\left. \sum_{t=1}^{J} P_{ts} g(\boldsymbol{x}_{ts}) \right) \right]. \qquad (5.2.10)
\end{aligned}$$

When $\boldsymbol{x}_d \in C_s$ we have

$$P_{js} = \frac{\exp\left(\frac{g(\boldsymbol{x}_{js})}{\sigma}\right)}{\sum_{t=1}^{J-1} \exp\left(\frac{g(\boldsymbol{x}_{ts})}{\sigma}\right)} \left\{1 - \exp\left[-\sum_{t=1}^{J-1} \exp\left(\frac{g(\boldsymbol{x}_{ts}) + \mu}{\sigma}\right)\right]\right\}, \tag{5.2.11}$$

and the derivatives of the choice probabilities are derived as

$$\begin{aligned}
\frac{\partial^2}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} \log(P_{js}) &= \frac{\partial}{\partial \boldsymbol{\beta}^T}\left[\frac{\partial}{\partial \boldsymbol{\beta}} \log(P_{js})\right], \\
&= \frac{\partial}{\partial \boldsymbol{\beta}^T}\left\{\frac{1}{\sigma}\left[-\boldsymbol{x}_{js}^T + \frac{\sum_{t=1}^{J-1} \exp\left(\frac{g(\boldsymbol{x}_{ts})}{\sigma}\right) \boldsymbol{x}_{ts}^T}{\sum_{t=1}^{J-1} \exp\left(\frac{g(\boldsymbol{x}_{ts})}{\sigma}\right)} - \right.\right. \\
&\quad \left.\left. \frac{\exp\left[-\sum_{t=1}^{J-1} \exp\left(\frac{g(\boldsymbol{x}_{ts})+\mu}{\sigma}\right)\right] \sum_{t=1}^{J-1} \exp\left[\frac{g(\boldsymbol{x}_{ts})+\mu}{\sigma}\right] \boldsymbol{x}_{ts}^T}{\left\{1 - \exp\left[-\sum_{t=1}^{J-1} \exp\left(\frac{g(\boldsymbol{x}_{ts})+\mu}{\sigma}\right)\right]\right\}}\right]\right\}, \\
&= \frac{\partial}{\partial \boldsymbol{\beta}^T}\left\{\frac{1}{\sigma}\left[-\boldsymbol{x}_{js}^T + \sum_{t=1}^{J-1} P_{ts} \boldsymbol{x}_{ts}^T - \frac{P_{Js} \sum_{t=1}^{J-1} \exp\left[\frac{g(\boldsymbol{x}_{ts})+\mu}{\sigma}\right] \boldsymbol{x}_{ts}^T}{1 - P_{Js}}\right]\right\}, \\
&= -\frac{1}{\sigma^2}\left[\sum_{t=1}^{J-1} P_{ts} \boldsymbol{x}_{ts}^T \boldsymbol{x}_{ts} - \sum_{t=1}^{J-1} P_{ts} \boldsymbol{x}_{ts}^T \sum_{t=1}^{J-1} P_{ts} \boldsymbol{x}_{ts} + \right. \\
&\quad \left. \frac{P_{Js}}{1 - P_{Js}} \sum_{t=1}^{J-1} \exp\left(\frac{g(\boldsymbol{x}_{ts}) + \mu}{\sigma}\right) K_{ts} \boldsymbol{x}_{ts}^T \boldsymbol{x}_{ts}\right], \\
&= -\frac{1}{\sigma^2}\left[\boldsymbol{X}_{s_{\{-d\}}} (\boldsymbol{P}_{s_{\{-d\}}} - \boldsymbol{p}_{s_{\{-d\}}} \boldsymbol{p}_{s_{\{-d\}}}^T) \boldsymbol{X}_{s_{\{-d\}}}^T + \right. \\
&\quad \left. \frac{P_{Js}}{1 - P_{Js}} \sum_{t=1}^{J-1} \exp\left(\frac{g(\boldsymbol{x}_{ts}) + \mu}{\sigma}\right) K_{ts} \boldsymbol{x}_{ts}^T \boldsymbol{x}_{ts}\right],
\end{aligned} \tag{5.2.12}$$

where $\mathbf{X}_{s_{\{-d\}}} = [\boldsymbol{x}_{1s}^T, \ldots, \boldsymbol{x}_{(J-1)s}^T]$ is the corresponding design matrix for choice set s excluding the representation of the death state, $\mathbf{P}_{s_{\{-d\}}}$ and $\mathbf{p}_{s_{\{-d\}}}$ are a $(J-1) \times (J-1)$ diagonal matrix and a $(J-1) \times 1$ vector with elements j, j and j both given by the choice probabilities for each health state $\boldsymbol{x}_{js}$ in a death choice set except the death

state, respectively, and

$$K_{ts} = \left\{ \sum_{t=1}^{J-1} \exp\left[\frac{g(\boldsymbol{x}_{ts}) + \mu}{\sigma} \right] - 1 + \frac{P_{Js}}{1 - P_{Js}} \sum_{t=1}^{J-1} \exp\left[\frac{g(\boldsymbol{x}_{ts}) + \mu}{\sigma} \right] \right\}.$$

Taking the derivative with respect to σ gives

$$\begin{aligned}
\frac{\partial^2}{\partial\sigma\partial\boldsymbol{\beta}} \log(P_{js}) &= \frac{\partial}{\partial\sigma} \left[\frac{\partial}{\partial\boldsymbol{\beta}} \log(P_{js}) \right], \\
&= \frac{\partial}{\partial\sigma} \left\{ \frac{1}{\sigma} \left[-\boldsymbol{x}_{js}^T + \sum_{t=1}^{J-1} P_{ts} \boldsymbol{x}_{ts}^T - \frac{P_{Js}}{1 - P_{Js}} \sum_{t=1}^{J-1} \exp\left(\frac{g(\boldsymbol{x}_{ts}) + \mu}{\sigma} \right) \boldsymbol{x}_{ts}^T \right] \right\}, \\
&= \frac{1}{\sigma^3} \left[\sigma(\boldsymbol{x}_{js}^T - \mathbf{X}_{s_{\{-d\}}} \mathbf{p}_{s_{\{-d\}}}) - \mathbf{X}_{s_{\{-d\}}} (\mathbf{P}_{s_{\{-d\}}} - \mathbf{p}_{s_{\{-d\}}} \mathbf{p}_{s_{\{-d\}}}^T) \mathbf{U}_{s_{\{-d\}}}^T + \right. \\
&\quad \left. \frac{P_{Js}}{1 - P_{Js}} F_{ts} \boldsymbol{x}_{ts}^T \right], \qquad (5.2.13)
\end{aligned}$$

where $\mathbf{X}_{s_{\{-d\}}}$, $\mathbf{P}_{s_{\{-d\}}}$, and $\mathbf{p}_{s_{\{-d\}}}$ as defined in Equation (5.2.12), and $\mathbf{U}_{s_{\{-d\}}}$ is a row vector of the corresponding mean utility values of the alternatives in choice set s excluding the mean utility value of death; and

$$\begin{aligned}
F_{ts} &= \sigma \sum_{t=1}^{J-1} \exp\left(\frac{g(\boldsymbol{x}_{ts}) + \mu}{\sigma} \right) + \sum_{t=1}^{J-1} (g(\boldsymbol{x}_{ts}) + \mu) \exp\left(\frac{g(\boldsymbol{x}_{ts}) + \mu}{\sigma} \right) \times \\
&\quad \left[1 - \frac{1}{1 - P_{Js}} \sum_{t=1}^{J} \exp\left(\frac{g(\boldsymbol{x}_{ts}) + \mu}{\sigma} \right) \right].
\end{aligned}$$

The second derivative with respect to σ is derived as

$$\begin{aligned}
\frac{\partial^2}{\partial\sigma^2}\log(P_{js}) &= \frac{\partial}{\partial\sigma}\left[\frac{\partial}{\partial\sigma}\log(P_{js})\right], \\
&= \frac{\partial}{\partial\sigma}\left\{\frac{1}{\sigma^2}\left[-g(\boldsymbol{x}_{js})+\sum_{t=1}^{J-1}P_{ts}g(\boldsymbol{x}_{ts})-\right.\right. \\
&\quad \left.\left.\frac{P_{Js}}{1-P_{Js}}\sum_{t=1}^{J-1}(g(\boldsymbol{x}_{ts})+\mu)\exp\left(\frac{g(\boldsymbol{x}_{ts})+\mu}{\sigma}\right)\right]\right\}, \\
&= \frac{1}{\sigma^4}\left\{2\sigma g(\boldsymbol{x}_{js})-\sum_{t=1}^{J-1}P_{ts}g(\boldsymbol{x}_{ts})\left[2\sigma+g(\boldsymbol{x}_{ts})-\sum_{t=1}^{J-1}P_{ts}g(\boldsymbol{x}_{ts})\right]+\right. \\
&\quad \frac{P_{Js}}{(1-P_{Js})}\sum_{t=1}^{J-1}(g(\boldsymbol{x}_{ts})+\mu)\exp\left(\frac{g(\boldsymbol{x}_{ts})+\mu}{\sigma}\right)[2\sigma+g(\boldsymbol{x}_{ts})+\mu- \\
&\quad \left.\frac{1}{1-P_{Js}}\sum_{t=1}^{J-1}\left(\frac{g(\boldsymbol{x}_{ts})+\mu}{\sigma}\right)\exp\left(\frac{g(\boldsymbol{x}_{ts})+\mu}{\sigma}\right)\right]\right\}. \qquad (5.2.14)
\end{aligned}$$

For the death state, where P_{Js} is the probability of death state,

$$\begin{aligned}
\frac{\partial^2}{\partial\boldsymbol{\beta}^2}\log(P_{Js}) &= \frac{\partial}{\partial\boldsymbol{\beta}^T}\left\{\frac{\partial}{\partial\boldsymbol{\beta}}\log\exp\left[-\sum_{t=1}^{J-1}\exp\left(\frac{g(\boldsymbol{x}_{it})+\mu}{\sigma}\right)\right]\right\}, \\
&= -\frac{1}{\sigma^2}\sum_{t=1}^{J-1}\exp\left(\frac{g(\boldsymbol{x}_{ts})+\mu}{\sigma}\right)\boldsymbol{x}_{ts}^T\boldsymbol{x}_{ts}, \qquad (5.2.15)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2}{\partial\sigma\partial\boldsymbol{\beta}}\log(P_{Js}) &= \frac{\partial}{\partial\sigma}\left\{\frac{\partial}{\partial\boldsymbol{\beta}}\log\exp\left[-\sum_{t=1}^{J-1}\exp\left(\frac{g(\boldsymbol{x}_{it})+\mu}{\sigma}\right)\right]\right\}, \\
&= -\frac{1}{\sigma^3}\left[\sigma\sum_{t=1}^{J-1}\exp\left(\frac{g(\boldsymbol{x}_{ts})+\mu}{\sigma}\right)+\right. \\
&\quad \left.\sum_{t=1}^{J-1}(g(\boldsymbol{x}_{ts})+\mu)\exp\left(\frac{g(\boldsymbol{x}_{ts})+\mu}{\sigma}\right)\right]\boldsymbol{x}_{ts}^T, \qquad (5.2.16)
\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2}{\partial\sigma^2}\log(P_{Js}) &= \frac{\partial}{\partial\sigma}\left\{\frac{\partial}{\partial\sigma}\log\exp\left[-\sum_{t=1}^{J-1}\exp\left(\frac{g(\boldsymbol{x}_{it})+\mu}{\sigma}\right)\right]\right\},\\ &= -\frac{1}{\sigma^4}\sum_{t=1}^{J-1}(g(\boldsymbol{x}_{ts})+\mu)\exp\left(\frac{g(\boldsymbol{x}_{ts})+\mu}{\sigma}\right)\times\\ &\quad [2\sigma+g(\boldsymbol{x}_{ts})+\mu]. \end{aligned} \tag{5.2.17}$$

The sub-FIMs of the partitioned FIM for the MNL model from one respondent defined in Equations (5.2.5), (5.2.6) and (5.2.7) are then computed by substituting the corresponding second derivative derived in Equations (5.2.8), (5.2.9), (5.2.10), (5.2.12), (5.2.13), (5.2.14), (5.2.15), (5.2.16) and (5.2.17), and then taking the negative expectation. Using the fact that responses are independent and each represent a random draw from a multinomial distribution, together with some simple algebra, then the partitioned FIM obtained for one respondent can be written as

$$\mathrm{FIM}_I(\xi,\boldsymbol{\theta}) = \begin{bmatrix}\sum_{s=1}^{S}\mathrm{FIM}_s^{11}(\xi,\boldsymbol{\theta}) & \sum_{s=1}^{S}\mathrm{FIM}_s^{12}(\xi,\boldsymbol{\theta})\\ \sum_{s=1}^{S}\mathrm{FIM}_s^{21}(\xi,\boldsymbol{\theta}) & \sum_{s=1}^{S}\mathrm{FIM}_s^{22}(\xi,\boldsymbol{\theta})\end{bmatrix},$$

where FIM_s is the Fisher information matrix obtained for choice set s, defined as

$$\mathrm{FIM}_s^{11}(\xi,\boldsymbol{\theta}) = \begin{cases}\frac{1}{\sigma^2}\left[\mathbf{X}_s(\mathbf{P}_s-\mathbf{p}_s\mathbf{p}_s^T)\mathbf{X}_s^T\right], & \text{if } \boldsymbol{x}_d\notin C_s;\\ \frac{1}{\sigma^2}\left[(1-P_{Js})\boldsymbol{X}_{s_{\{-d\}}}\left(\boldsymbol{P}_{s_{\{-d\}}}-\boldsymbol{p}_{s_{\{-d\}}}\boldsymbol{p}_{s_{\{-d\}}}^T\right)\boldsymbol{X}_{s_{\{-d\}}}^T+\right.\\ \left.\frac{P_{Js}}{(1-P_{Js})}\sum_{t=1}^{J-1}\left[\exp\left(\frac{g(\boldsymbol{x}_{ts})+\mu}{\sigma}\right)\right]^2\boldsymbol{x}_{ts}^T\boldsymbol{x}_{ts}\right], & \text{if } \boldsymbol{x}_d\in C_s.\end{cases}$$

$$\text{FIM}_s^{12}(\xi,\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^3}\left[\mathbf{X}_s(\mathbf{P}_s - \mathbf{p}_s\mathbf{p}_s^T)\mathbf{U}_s^T\right], & \text{if } \boldsymbol{x}_d \notin C_s; \\ \frac{1}{\sigma^3}\left\{(1-P_{Js})\left[\mathbf{X}_{s_{\{-d\}}}(\mathbf{P}_{s_{\{-d\}}} - \mathbf{p}_{s_{\{-d\}}}\mathbf{p}_{s_{\{-d\}}}^T)\mathbf{U}_{s_{\{-d\}}}^T\right] + \right. & \\ \frac{P_{Js}}{1-P_{Js}}\left[\sum_{t=1}^{J-1}(g(\boldsymbol{x}_{ts})+\mu)\exp\left(\frac{g(\boldsymbol{x}_{ts})+\mu}{\sigma}\right)\sum_{t=1}^{J-1}\exp\left(\frac{g(\boldsymbol{x}_{ts})+\mu}{\sigma}\right)\times\right. & \\ \left.\left.\sum_{t=1}^{J-1}\exp\left(\frac{g(\boldsymbol{x}_{ts})+\mu}{\sigma}\right)\right]\boldsymbol{x}_{ts}^T - \sigma P_{Js}\mathbf{X}_{s_{\{-d\}}}\mathbf{p}_{s_{\{-d\}}}\right\}, & \text{if } \boldsymbol{x}_d \in C_s. \end{cases}$$

and

$$\text{FIM}_s^{22}(\xi,\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^4}\left[\sum_{t=1}^{J} g(\boldsymbol{x}_{ts})^2 P_{ts} - \left(\sum_{t=1}^{J} g(\boldsymbol{x}_{ts})P_{ts}\right)^2\right], & \text{if } \boldsymbol{x}_d \notin C_s; \\ \frac{1}{\sigma^4}\left\{(1-P_{Js})\left[\sum_{t=1}^{J-1} g(\boldsymbol{x}_{ts})^2 P_{ts} - \left(\sum_{t=1}^{J-1} g(\boldsymbol{x}_{ts})P_{ts}\right)^3\right] + \right. & \\ \frac{P_{Js}}{1-P_{Js}}\left[(g(\boldsymbol{x}_{ts})+\mu)\exp\left(\frac{g(\boldsymbol{x}_{ts})+\mu}{\sigma}\right)\right]^2 - & \\ \left.\sigma P_{Js}\sum_{t=1}^{J-1} P_{ts}g(\boldsymbol{x}_{ts})\right\}, & \text{if } \boldsymbol{x}_d \in C_s. \end{cases}$$

Hence, the FIM from N independent respondents becomes

$$\begin{aligned} \text{FIM}(\xi,\boldsymbol{\theta}) &= N\ \text{FIM}_I(\xi,\boldsymbol{\theta}) \\ &= N\begin{bmatrix} \sum_{s=1}^{S}\text{FIM}_s^{11}(\xi,\boldsymbol{\theta}) & \sum_{s=1}^{S}\text{FIM}_s^{12}(\xi,\boldsymbol{\theta}) \\ \sum_{s=1}^{S}\text{FIM}_s^{21}(\xi,\boldsymbol{\theta}) & \sum_{s=1}^{S}\text{FIM}_s^{22}(\xi,\boldsymbol{\theta}) \end{bmatrix}. \end{aligned} \tag{5.2.18}$$

The derivation of the Bayesian $\mathcal{D}_{\mathbf{S}}$-optimal design requires minimising the volume of the credible interval of the parameter vector $\boldsymbol{\beta}$, which is asymptotically equivalent to minimising the determinant of the inverse of the $\text{FIM}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta})$ after substituting the

definition of the sub-FIMs in Equation (5.2.4), that gives

$$\begin{aligned}
\mathrm{FIM}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta}) &= \mathrm{FIM}^{11}(\xi,\boldsymbol{\theta}) - \mathrm{FIM}^{12}(\xi,\boldsymbol{\theta})\left[\mathrm{FIM}^{22}(\xi,\boldsymbol{\theta})\right]^{-1}\mathrm{FIM}^{21}(\xi,\boldsymbol{\theta}), \\
&= N\left\{\sum_{s=1}^{S}\mathrm{FIM}_s^{11}(\xi,\boldsymbol{\theta}) - \sum_{s=1}^{S}\mathrm{FIM}_s^{12}(\xi,\boldsymbol{\theta})\left[\sum_{s=1}^{S}\mathrm{FIM}_s^{22}(\xi,\boldsymbol{\theta})\right]^{-1}\times\right. \\
&\quad \left.\left[\sum_{s=1}^{S}\mathrm{FIM}_s^{12}(\xi,\boldsymbol{\theta})\right]^{T}\right\} \qquad (5.2.19)
\end{aligned}$$

However, this information matrix depends on the value of the unknown parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)$ through the choice probabilities and the mean utility values, and hence so does the construction of the Bayesian design. In the Bayesian approach, we average the value of the design criterion over a suitable prior distribution for the unknown model parameters, $\pi(\boldsymbol{\theta})$. Formally, the asymptotic Bayesian $\mathcal{D}_{\mathbf{S}}$-optimality criterion based on the FIM for the MNL model can be defined as

$$\mathcal{D}^{B}_{\mathbf{S},\mathrm{FIM}} = \int \left|\mathrm{FIM}^{-1}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta})\right| \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}, \qquad (5.2.20)$$

where k is the total number of the unknown model parameters. We raise the determinant to the power of $1/k$ to standardise the statistical measure and make irrelevant to the dimension of the model.

An alternative for the FIM approximation, particularly with small sample sizes, is the GFIM. This approximation, as mentioned in Section 3.6.1, has better finite sample properties than the FIM, is a better approximation for the posterior VCM of the model parameters, and hence might result in better choice design. The asymptotic Bayesian $\mathcal{D}_{\mathbf{S}}$-optimality criterion based on the GFIM for the MNL model can be formulated as

$$\mathcal{D}^{B}_{\mathbf{S},\mathrm{GFIM}} = \int \left|\mathrm{GFIM}^{-1}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta})\right|^{1/k} \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}, \qquad (5.2.21)$$

where

$$\begin{aligned}\text{GFIM}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta}) &= N\left\{\sum_{s=1}^{S}\text{GFIM}_s^{11}(\xi,\boldsymbol{\theta})-\sum_{s=1}^{S}\text{GFIM}_s^{12}(\xi,\boldsymbol{\theta})\left[\sum_{s=1}^{S}\text{GFIM}_s^{22}(\xi,\boldsymbol{\theta})\right]^{-1}\times\right.\\ &\left.\left[\sum_{s=1}^{S}\text{GFIM}_s^{12}(\xi,\boldsymbol{\theta})\right]^{T}\right\} \qquad (5.2.22)\end{aligned}$$

and the sub-matrices of the partitioned GFIM can be obtained from the partitioned FIM as

$$\begin{aligned}\text{GFIM}(\xi,\boldsymbol{\theta}) &= \begin{bmatrix}\sum_{s=1}^{S}\text{GFIM}_s^{11}(\xi,\boldsymbol{\theta}) & \sum_{s=1}^{S}\text{GFIM}_s^{12}(\xi,\boldsymbol{\theta})\\ \sum_{s=1}^{S}\text{GFIM}_s^{21}(\xi,\boldsymbol{\theta}) & \sum_{s=1}^{S}\text{GFIM}_s^{22}(\xi,\boldsymbol{\theta})\end{bmatrix},\\ &= \text{FIM}(\xi,\boldsymbol{\theta})+\Sigma^{-1} \qquad (5.2.23)\end{aligned}$$

where Σ^{-1} is the inverse of the prior VCM that represents the amount of information a prior carries about the unknown model parameters, $\boldsymbol{\theta}$.

The design criteria derived above are for multinomial choice experiments. However, in this thesis, we only consider pairwise choice experiments. We simplify the notation by setting $J=2$. Thus, assuming that, for choice sets including death, the death state is the second alternative, the sub-FIMs can be simplified to

$$\text{FIM}_s^{11}(\xi,\boldsymbol{\theta})=\begin{cases}\frac{1}{\sigma^2}(\boldsymbol{x}_{1s}-\boldsymbol{x}_{2s})^T P_{1s}(1-P_{1s})(\boldsymbol{x}_{1s}-\boldsymbol{x}_{2s}), & \text{if } \boldsymbol{x}_d\notin C_s;\\ \\ \frac{1}{\sigma^2}\frac{P_{Js}}{(1-P_{Js})}\left[\exp\left(\frac{g(\boldsymbol{x}_{1s})+\mu}{\sigma}\right)\right]^2\boldsymbol{x}_{1s}^T\boldsymbol{x}_{1s}, & \text{if } \boldsymbol{x}_d\in C_s,\end{cases} \qquad (5.2.24)$$

$$\text{FIM}_s^{12}(\xi,\boldsymbol{\theta})=\begin{cases}\frac{1}{\sigma^3}(\boldsymbol{x}_{1s}-\boldsymbol{x}_{2s})^T P_{1s}(1-P_{1s})\left[g(\boldsymbol{x}_{1s})-g(\boldsymbol{x}_{2s})\right], & \text{if } \boldsymbol{x}_d\notin C_s;\\ \\ \frac{1}{\sigma^3}\left\{P_{1s}^2(1-P_{1s})g(\boldsymbol{x}_{1s})-\sigma P_{1s}(1-P_{1s})+\right. & \\ \left.\frac{P_{Js}}{1-P_{Js}}(g(\boldsymbol{x}_{1s})+\mu)\left[\exp\left(\frac{g(\boldsymbol{x}_{1s})+\mu}{\sigma}\right)\right]^2\right\}\boldsymbol{x}_{1s}^T, & \text{if } \boldsymbol{x}_d\in C_s,\end{cases} \qquad (5.2.25)$$

and

$$\text{FIM}_s^{22}(\xi,\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^4} P_{1s}(1-P_{1s})\left[g(\boldsymbol{x}_{1s}) - g(\boldsymbol{x}_{2s})\right]^2, & \text{if } \boldsymbol{x}_d \notin C_s; \\ \frac{1}{\sigma^4}\left\{P_{1s}[g(\boldsymbol{x}_{1s})^2 P_{1s} - g(\boldsymbol{x}_{1s})^3 P_{1s}^3] - 2\sigma P_{1s}(1-P_{1s}) g(\boldsymbol{x}_{1s}) + \right. & \\ \left. \frac{P_{JS}}{1-P_{JS}}\left[(g(\boldsymbol{x}_{1s})+\mu)\exp\left(\frac{g(\boldsymbol{x}_{1s})+\mu}{\sigma}\right)\right]^2\right\}, & \text{if } \boldsymbol{x}_d \in C_s, \end{cases} \tag{5.2.26}$$

and $(\boldsymbol{x}_{1s} - \boldsymbol{x}_{2s})$ is the difference between attribute levels of first and second alternatives presented in choice set s, P_{1s} and P_{Js} are the choice probabilities of the first alternative in each choice question and the death state, respectively. The asymptotic Bayesian $\mathcal{D}_{\mathbf{S}}$-optimality criteria based on FIM and GFIM for the binomial logit model, which is also called the logit model, is then obtained by substituting these sub-FIM$_s$ defined in Equations (5.2.24) and (5.2.25) and (5.2.26) in Equations (5.2.19) and (5.2.22), respectively.

5.3

Constructing Choice Design Using Available Software

In this section, we investigate the ability of Bayesian choice design algorithms available in software such as SAS, JMP (Kessels, 2010) and Ngene (Rose and Bliemer, 2012) to generate a pairwise experiment for valuing health states while taking into account our design considerations. In particular, we consider their ability to

- allow the death state to appear automatically in the choice design, hence optimising the correct design criterion which accounts for the inclusion of death state

in the MNL model;

- set up design constraints to avoid implausible attribute combinations and dominant choice sets, to improve the efficiency of the choice design;
- specify different prior distributions for the unknown model parameters, $\boldsymbol{\beta}$ and σ, to construct Bayesian choice designs, since we aim to investigate the effect of the choice of prior on the choice of the optimal design itself.

We start by looking at the properties of the software in terms of DCMs covered, design criteria, and algorithms used to construct choice sets, in addition to other options provided by the software as shown in Table 5.1.

We found that Ngene was the most flexible software in terms of having variety of DCMs, Bayesian design criteria, design algorithms and prior distributions, as illustrated in Table 5.1, as well as other design options such as the possibility of having a different definition for the utility function and including the death state as a common choice. However, we have not managed to construct Bayesian choice designs that totally handle our design problem, particularly in terms of including the death state automatically in the choice design and hence optimising the correct design criterion. This is because the design software is built generally to construct choice designs in any field and not specifically for health economic evaluation studies. Therefore, they do not necessarily cover the specific requirements for generating choice designs for valuing health states, particularly in terms of including the death state to anchor health state utility values. The general limitations of the design software can be summarised as follows.

1. **Design Criterion**

   Our design problem is to construct Bayesian optimal pairwise choice designs that minimise the variance of the preference parameters, $\boldsymbol{\beta}$, and hence the variance of the mean health state utilities estimated within the QALY scale. Following our method of anchoring health state utility into the QALY scale, by including

Table 5.1: Design software and algorithms to construct Bayesian optimal choice design

Software	***DCM***	***Design Criterion***	***Prior Distribution***	***Design Algorithm***	***User's Constraints***
SAS/ChoicEFF macro	MNL	$\mathcal{D}_B$	Multivariate normal	MCMF (Kessels et al., 2004)	Specifying candidate alternatives and choice set using MKTEX macro
SAS/JMP	MNL	$\mathcal{D}_B$ and $\mathcal{V}$	Multivariate normal	Coordinate-exchange algorithm (Meyer and Nachtsheim, 1995)	Specify constraints using custom design
MATLAB/Stat toolbox	MNL	$\mathcal{D}_B$, $\mathcal{A}_B$, $\mathbf{G}_B$ and $\mathcal{V}_B$	Multivariate normal	Adaptive algorithm (Kessels et al., 2009)	Specify design constraints using design editor that has four constraint types
Ngene	MNL, Mixed MNL, Cross sectional MNL and Panel Mixed MNL	$\mathcal{D}_B$, $\mathcal{A}_B$, and $\mathcal{S}_B$	Variety of prior distributions	MCMF (Kessels et al., 2004), RSC algorithm (Sándor and Wedel, 2001), Swapping algorithm (Rose and Bliemer, 2008)	Specify constraints using conditional statements

the death state in the choice design, this requires optimising the Bayesian $\mathcal{D}_{\mathbf{S}}$ optimality criterion

$$\mathcal{D}^{B}_{\mathbf{S},\text{FIM}} = \int \left|\text{FIM}^{-1}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta})\right| \pi(\boldsymbol{\theta}) d\boldsymbol{\theta},$$

where

$$\begin{aligned}
\text{FIM}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta}) &= \text{FIM}^{11}(\xi,\boldsymbol{\theta}) - \text{FIM}^{12}(\xi,\boldsymbol{\theta})\left[\text{FIM}^{22}(\xi,\boldsymbol{\theta})\right]^{-1}\text{FIM}^{21}(\xi,\boldsymbol{\theta}), \\
&= N\left\{\sum_{s=1}^{S}\text{FIM}^{11}_{s}(\xi,\boldsymbol{\theta}) - \sum_{s=1}^{S}\text{FIM}^{12}_{s}(\xi,\boldsymbol{\theta})\left[\sum_{s=1}^{S}\text{FIM}^{22}_{s}(\xi,\boldsymbol{\theta})\right]^{-1}\times \right. \\
&\quad \left.\left[\sum_{s=1}^{S}\text{FIM}^{12}_{s}(\xi,\boldsymbol{\theta})\right]^{T}\right\},
\end{aligned}$$

and the sub-matrices of the partitioned FIM are defined as in Equations (5.2.24) and (5.2.25) and (5.2.26).

However, the software does not allow us to optimise the choice design with respect to the $\mathcal{D}_{\mathbf{S}}$-optimality criterion defined above, as the provided $\mathcal{D}$-optimality criterion do not account for the addition of the death state in the choice design which is represented by, for example, $\frac{1}{\sigma^2}\frac{P_{Js}}{(1-P_{Js})}\left[\exp\left(\frac{g(\boldsymbol{x}_{1s})+\mu}{\sigma}\right)\right]^2\boldsymbol{x}^{T}_{1s}\boldsymbol{x}_{1s}$ in the $\text{FIM}^{11}_{s}(\xi,\boldsymbol{\theta})$. This is because the design criterion in the software is derived for the general MNL model, which does not account for the addition of death, and this results in a slightly different design criterion.

Additionally, the softwares use Bayesian $\mathcal{D}$-optimality criterion, i.e. seek designs that minimise the posterior credible ellipsoid of the MNL model parameters, whose approximation is based on the FIM. In this approximation, the posterior VCM of the preference parameters is approximated using the inverse of the FIM of the parameters treating the scale parameter in the MNL, σ, as a fixed parameter. However, this thesis considers calculating Bayesian optimal paired comparison design for the logit model considering the preference parameters $\boldsymbol{\beta}$ and assuming the scale parameter σ is unknown nuisance parameter, which re-

sults in slightly different FIM and, hence, design criterion. Also, we consider the GFIM approximation of the VCM of the MNL model parameters, that result in the $\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$ design criterion, in order to study the effect of incorporating the prior information in the approximation of the posterior VCM in the choice of the optimal design. This approximation is known to be a better approximation for the posterior VCM of the unknown model parameters (Yu et al., 2008), which might lead to more efficient choice designs. This cannot be investigated using the available software.

2. **Prior Distribution**

   All the design softwares, except Ngene, assume a multivariate normal distribution for the unknown model parameters to generate a Bayesian choice design. In our design problem, the values of the parameters should be positive, as each represents the incremental decrease in mean utility when moving one level in one attribute. This distribution is not appropriate as it produces negative values, except under some very restrictive conditions on the mean and the variance. Hence, other prior distributions with zero probability for negative values are required, such as Gamma and Beta distributions.

3. **Design Software Modification**

   Design software such as JMP and Ngene are not available for all users, as they are commercial software. Therefore, modifying design models and criteria specified in these softwares to handle a specific design problem is not an easy task and typically requires a higher upfront cost to purchase the application.

Given these limitations, we consider the need for deriving a efficient methodology to generate Bayesian optimal choice design for provision of health state utilities within the required QALY scale and considering the correct design criterion. This can be done by modifying the advanced choice design algorithms introduced in Section 3.7.2 and used in some of these software to handle our specific design applications.

5.4 Bayesian Design Algorithm for Generating DCE for Valuing Health State Utilities

In this section, we describe our design algorithm that we used to generate efficient Bayesian paired comparison designs, particulary for pairwise experiments, based on the logit model and Bayesian design criteria defined in Section 5.2. Initially, we produce a computer search algorithm that relies on a random search through a reasonably large number of choice designs, each with the required number of alternatives and choice questions, as will be described in Section 5.4.1. To further improve the random design, we use a more efficient search algorithm, namely the coordinate-exchange algorithm by Meyer and Nachtsheim (1995), together with an updating formula for the information matrix to accelerate the computational time of Bayesian $\mathcal{D}_{\mathbf{S}}$ optimal design as described in Sections 5.4.2 and 5.4.3, respectively. The algorithm is programmed in the R language (R Core Team, 2013), and applied to construct Bayesian pairwise choice design for the AQL-5D case study in Section 5.6.

5.4.1 Random Search Algorithm

The random search algorithm is basically an iterative search procedure based on a random search over a large number of designs. These designs are based on the desirable number of choice sets, pairwise comparisons. Each satisfies our design constraints of excluding dominant and implausible options as well as including the death state. The

procedure of this algorithm can be described in the following three steps.

1. We generate a reasonable number of choice designs, each with the required number of pairs, where at least one of these pairs is a death comparison. For each design, the alternatives in non-death pairs were selected randomly, but with constraints to prevent implausible and dominant alternatives from appearing in the final choice design. The algorithm identifies and excludes these alternatives during the construction of each choice design based on the attribute levels of each alternative and using conditional statements in the R program to define our constraints on the attribute levels.

   For the AQL-5D classification system, a health state $\boldsymbol{x}_{ijs}$ is more likely to be unrealistic if it has most attributes levels at very mild level (e.g. at levels 0 and 1) with the remaining attributes at very severe levels (e.g. levels 3 and 4) or vice versa. An example of an implausible health state is the AQL-5D health state 00034 where a person has no concern about having asthma and no problem with breathing or the weather condition while suffering from an extreme limitation on all activities done. Therefore, in our design we define the state as unrealistic if the sum of the first three attributes' levels is less than two while the sum of last two attributes is greater than four. In terms of the definition of the health state $\boldsymbol{x}_{ijs}$,this can be defined as in Equation (3.3.1)

$$\sum_{\lambda=1}^{4}\sum_{\delta=1}^{3} x_{\lambda\delta} < 2 \quad \& \quad \sum_{\lambda=1}^{4}\sum_{\delta=4}^{5} x_{\lambda\delta} \geqslant 4.$$

   Therefore, in our algorithm when generating the alternatives for any pair in the choice design, if the sum of the first three attributes of an alternative is less than two, then the algorithm excludes the option of selecting very severe levels, such as levels 3 and 4, for the last two attributes, sleep and activities and vice versa. For pairwise comparison with dominant option (e.g. AQL-5D health state 02111 that dominates 02123), the algorithm generates another alternative for that pair.

In death pairwise comparisons, the alternative for the death state is selected such that it could be compared to the death state and respondents are able to make trade-off between both states. Therefore, the alternative for the death state is usually represented by the more severe health states defined by the underlying classification system such as the AQL-5D health states 44444 and 44434.

2. We then compare the generated random designs based on their Bayesian $\mathcal{D}_{\mathbf{S}}$ criterion values computed as

$$\begin{aligned}
\mathcal{D}^B_{\mathbf{S},\text{FIM}} &= \int \left|\text{FIM}^{-1}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta})\right| \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}, \\
&= \int \left| \frac{1}{N} \left\{ \sum_{s=1}^{S} \text{FIM}^{11}_s(\xi,\boldsymbol{\theta}) - \sum_{s=1}^{S} \text{FIM}^{12}_s(\xi,\boldsymbol{\theta}) \times \right.\right. \\
&\quad \left.\left. \left[\sum_{s=1}^{S} \text{FIM}^{22}_s(\xi,\boldsymbol{\theta})\right]^{-1} \left[\sum_{s=1}^{S} \text{FIM}^{12}_s(\xi,\boldsymbol{\theta})\right]^T \right\}^{-1} \right|^{1/k} \pi(\boldsymbol{\theta}) d\boldsymbol{\theta},
\end{aligned}$$

or when GFIM is used, as

$$\begin{aligned}
\mathcal{D}^B_{\mathbf{S},\text{GFIM}} &= \int \left|\text{GFIM}^{-1}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta})\right| \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}, \\
&= \int \left| \frac{1}{N} \left\{ \sum_{s=1}^{S} \text{GFIM}^{11}_s(\xi,\boldsymbol{\theta}) - \sum_{s=1}^{S} \text{GFIM}^{12}_s(\xi,\boldsymbol{\theta}) \times \right.\right. \\
&\quad \left.\left. \left[\sum_{s=1}^{S} \text{GFIM}^{22}_s(\xi,\boldsymbol{\theta})\right]^{-1} \left[\sum_{s=1}^{S} \text{GFIM}^{12}_s(\xi,\boldsymbol{\theta})\right]^T \right\}^{-1} \right|^{1/k} \pi(\boldsymbol{\theta}) d\boldsymbol{\theta},
\end{aligned}$$

where

$$\text{FIM}^{11}_s(\xi,\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^2}(\boldsymbol{x}_{1s} - \boldsymbol{x}_{2s})^T P_{1s}(1-P_{1s})(\boldsymbol{x}_{1s} - \boldsymbol{x}_{2s}), & \text{if } \boldsymbol{x}_d \notin C_s; \\ \\ \frac{1}{\sigma^2} \frac{P_{Js}}{(1-P_{Js})} \left[\exp\left(\frac{g(\boldsymbol{x}_{1s})+\mu}{\sigma}\right)\right]^2 \boldsymbol{x}^T_{1s}\boldsymbol{x}_{1s}, & \text{if } \boldsymbol{x}_d \in C_s, \end{cases}$$

$$\text{FIM}_s^{12}(\xi,\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^3}(\boldsymbol{x}_{1s}-\boldsymbol{x}_{2s})^T P_{1s}(1-P_{1s})\left[g(\boldsymbol{x}_{1s})-g(\boldsymbol{x}_{2s})\right], & \text{if } \boldsymbol{x}_d \notin C_s; \\ \frac{1}{\sigma^3}\left\{P_{1s}^2(1-P_{1s})g(\boldsymbol{x}_{1s}) - \sigma P_{1s}(1-P_{1s}) + \right. & \\ \left. \frac{P_{Js}}{1-P_{Js}}(g(\boldsymbol{x}_{1s})+\mu)\left[\exp\left(\frac{g(\boldsymbol{x}_{1s})+\mu}{\sigma}\right)\right]^2\right\}\boldsymbol{x}_{1s}^T, & \text{if } \boldsymbol{x}_d \in C_s, \end{cases}$$

and

$$\text{FIM}_s^{22}(\xi,\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^4}P_{1s}(1-P_{1s})\left[g(\boldsymbol{x}_{1s})-g(\boldsymbol{x}_{2s})\right]^2, & \text{if } \boldsymbol{x}_d \notin C_s; \\ \frac{1}{\sigma^4}\left\{P_{1s}[g(\boldsymbol{x}_{1s})^2 P_{1s} - g(\boldsymbol{x}_{1s})^3 P_{1s}^3] - 2\sigma P_{1s}(1-P_{1s})g(\boldsymbol{x}_{1s}) + \right. & \\ \left. \frac{P_{JS}}{1-P_{JS}}\left[(g(\boldsymbol{x}_{1s})+\mu)\exp\left(\frac{g(\boldsymbol{x}_{1s})+\mu}{\sigma}\right)\right]^2\right\}, & \text{if } \boldsymbol{x}_d \in C_s, \end{cases}$$

and the sub-matrices of the partitioned GFIM are obtained from the sub-FIMs of the partitioned FIM as defined in Equation (5.2.23).

The calculation of the Bayesian design criterion involves an integral that cannot be computed analytically. Therefore, we use Monte Carlo simulation to estimate the expected value of the design criterion over the prior distribution chosen for the model parameters. Thus, using R independent draws from the underlying prior distributions for each of the unknown logit model parameters, the optimality criterion value is approximated as

$$\hat{\mathcal{D}}_{\mathbf{S},\text{FIM}}^{B} = \frac{1}{R}\sum_{r=1}^{R}\mathcal{D}_{\mathbf{S},\text{FIM}}^{r}, \tag{5.4.1}$$

where $\mathcal{D}_{\mathbf{S},\text{FIM}}^{r}$ is the design criterion value computed at the r^{th} draw from the prior distributions. A similar approximation is derived for the asymptotic Bayesian criterion based on the GFIM, where

$$\hat{\mathcal{D}}_{\mathbf{S},\text{GFIM}}^{B} = \frac{1}{R}\sum_{r=1}^{R}\mathcal{D}_{\mathbf{S},\text{GFIM}}^{r}. \tag{5.4.2}$$

There are different ways of generating the draws from the given prior distributions for the model parameters; pseudo random sampling method is often used. In

this thesis the Bayesian design criterion values are approximated using the Latin hypercube sampling (LHS) instead of using simple random sampling, which is also known as the simple Monte Carlo (SMC) method. The reliability of the estimated value of the Bayesian $\mathcal{D}_{\mathbf{S}}$-optimality criterion depends on the variation in the random sample used, where the SMC requires larger number of draws to reduce the variation in the sample than the LHS. The LHS reduces the variation in the sample, and hence the variance of the estimator, by using a more systematic technique in generating the random sample from the prior distribution of each parameter, whereas the SMC method directly generates random draws from the required prior distribution. For a vector of k independent parameters, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$, we use the simple LHS technique that produces a random sample of size R for $\boldsymbol{\theta}$ by:

(a) dividing the sample space of θ_i into R regions of equal probability, $\frac{1}{R}$;

(b) randomly sampling a value from each region to obtain $\{\theta_{i,1}, \ldots, \theta_{i,R}\}$;

(c) randomly permuting the resulting R draws for θ_i to obtain $\{\theta^*_{i,1}, \ldots, \theta^*_{i,R}\}$; and

(d) combining the resulting r^{th} draw for each parameter $\theta_i \quad \text{for} \quad i = 1, \ldots, k$ to obtain the r^{th} LHS random sample for $\boldsymbol{\theta}$ as $(\theta^*_{1,r}, \ldots, \theta^*_{k,r})$.

Bliemer et al. (2008) showed that the SMC method performs badly in approximating the Bayesian optimality criterion value in comparison with other sampling methods such as LHS. Also, they showed that LHS performs equally well in terms of other advanced sampling methods such as the Halton and Sobol sequence sampling methods. The authors concluded that designs generated using the SMC method are less likely to be truly efficient, unless they were constructed using a substantially large number of random draws.

3. The algorithm returns the choice design with the smallest Bayesian $\mathcal{D}_{\mathbf{S}}$ criterion value. This design is selected out of 1,000 random choice designs, each evaluated using a sufficient number of LHS draws from the given prior distributions for

the unknown model parameters. The determination of the sufficient number of the LHS draws is based on the convergence of the criterion values, which will be considered in Section 5.6 when constructing Bayesian choice designs for the AQL-5D classification system.

The best random Bayesian optimal choice design returned by this procedure might not be the optimal choice design, as this depends on the number of iterations or random designs used. This can be investigated by running the algorithms for large numbers of designs (more than 1,000), which could take a long time, particularly for a large design problem. Therefore, the best random design produced by this random search procedure could be improved instead using a more advanced search algorithm, namely the coordinate-exchange algorithm developed in Meyer and Nachtsheim (1995). This algorithm is used in many software programs, such as JMP and Adaptive algorithm in MATLAB, to construct Bayesian optimal choice designs. The algorithm searches for the best attribute levels for each alternative in the choice design that optimises the underlying criterion value, as described in the following section.

5.4.2 Coordinate-exchange Algorithm

The coordinate-exchange algorithm is used here to improve the best random choice design resulting from the random search procedure. The algorithm is a column-based exchange algorithm that changes one attribute level of an alternative in a starting design (e.g. the best random design) at a time, and replaces it with the best exchange that improves the criterion value. The procedure of this algorithm can be summarised as follows.

1. Start with the first attribute for the first alternative in the starting design, the best random design, and cycle its level over all possible options (e.g. for an attribute with 5 levels, if the attribute is at level 2 then it is cycled over levels 3,4,5, and 1).

2. Compute the criterion value after each change, and execute the one that improves the Bayesian $\mathcal{D}_{\mathbf{S}}$ criterion value.

3. Pass the procedure through all the alternatives and attributes in the choice design until all exchanges have been investigated.

4. Start over again and repeat step 1-3 until no further possible improvement in the criterion value.

5. Return the design with the corresponding attribute levels that produce the best value for the underlying Bayesian optimality criterion.

The coordinate-exchange algorithm is usually run for many random starting designs to find the global or near global optimal designs. In our design algorithm, however, we will consider using 15 different starting designs to find the optimal or near optimal Bayesian design. This is because each of these starting designs is selected out of 1,000 Bayesian random designs using our random search algorithm, which is considerably time-consuming. Also, using the best Bayesian random design is expected to require less number of different starting design to find the global or near global Bayesian optimal design.

In addition, we modify this algorithm to account for the inclusion of the death state in the choice design, as well as the requirements of excluding unrealistic alternatives produced by the underlying classification system. We coded this algorithm such that it proceeds over all choice sets in the designs except the death choice sets, in order to avoid altering alternatives compared to the death state. In addition, to avoid producing unrealistic health states by the exchange procedure, we coded the algorithm such that it identifies all possible level exchanges for each attribute in each alternative while excluding attribute levels more likely to turn the underlying health state to implausible state. For instance, the possible level exchanges for the first attribute of the AQL-5D health state 31043 would be 4, 2, and 1, where level 0 is excluded as it turns the health state to an unrealistic alternative, 01043, by the given definition for the unrealistic

AQL-5D health state in Section 5.4.1.

The selection of the coordinate-exchange algorithm is based on the fact that it dramatically reduces the computational time for finding the best Bayesian choice design that optimises the criterion value compared to other choice design algorithms, particularly for large design problems, as shown in Kessels et al. (2009), who show that the coordinate-exchange algorithm is much more faster than the Monte Carlo Modified Fedorov (MCMF) algorithm developed in Kessels et al. (2004). This is because the former, as opposed to the MCMF algorithm, does not require a candidate set of all possible health states to do the exchange which grows exponentially with the number of attributes and attribute levels. For a large design problem, i.e. a design with a large number of attributes and attribute levels, the coordinate-exchange algorithm is more effective than the MCMF algorithm.

To further accelerate the computational time of the Bayesian $\mathcal{D}$ optimality criterion, Kessels et al. (2009) use an update formula for the FIM and the Cholesky decomposition of the updated FIM to economically compute the Bayesian criterion value for each possible change made by the coordinate-exchange algorithm. We use similar idea as in Kessels et al. (2009) in our design algorithm to economically compute the underlying Bayesian $\mathcal{D}_{\mathbf{S}}$ criterion value. In the following section, we derive the update formula for the FIM and GFIM after each change for our design model, and show how this formula is used to compute the $\mathcal{D}_{\mathbf{S}}$ criterion value using the Cholesky decomposition.

5.4.3 Updating the Information Matrix and the Cholesky Decomposition

The coordinate-exchange algorithm changes one attribute level of an alternative at a time. Hence the modified design, ξ^*, differs only in one alternative from the starting design, ξ. This means that the FIM of the modified design can be obtained by updating the FIM of the starting design according to this change, which we denote as ξ^*.

Since the coordinate-exchange algorithm in our pairwise experiment is applied over non-death pairs only, and following a similar idea in Meyer and Nachtsheim (1995) who use the coordinate-exchange algorithm for linear optimality, for each profile change the FIM* can be computed by adding and deleting the corresponding information matrices for the deleted and added alternatives in a particular non-death pair, denoted as FIM_s^- and FIM_s^+ respectively, defined for N respondents as

$$\begin{aligned}
\text{FIM}_s^{11}(\xi,\boldsymbol{\theta}) &= \frac{N}{\sigma^2}(\boldsymbol{x}_{1s}-\boldsymbol{x}_{2s})^T P_{1s}(1-P_{1s})(\boldsymbol{x}_{1s}-\boldsymbol{x}_{2s}),\\
\text{FIM}_s^{12}(\xi,\boldsymbol{\theta}) &= \frac{N}{\sigma^3}(\boldsymbol{x}_{1s}-\boldsymbol{x}_{2s})^T P_{1s}(1-P_{1s})\left[g(\boldsymbol{x}_{1s})-g(\boldsymbol{x}_{2s})\right],\\
\text{FIM}_s^{22}(\xi,\boldsymbol{\theta}) &= \frac{N}{\sigma^4}P_{1s}(1-P_{1s})\left[g(\boldsymbol{x}_{1s})-g(\boldsymbol{x}_{2s})\right]^2.
\end{aligned}$$

For illustration, suppose the first health state $\boldsymbol{x}_{1s}$ in a non-death pairwise comparison s is replaced by $\boldsymbol{x}_{1s}^+$, where the health state $\boldsymbol{x}_{1s}^+$ differs only in one attribute from the deleted health state, which we denote as $\boldsymbol{x}_{1s}^-$. The second alternative $\boldsymbol{x}_{2s}$ remains fixed, and hence the corresponding information matrix given this change can be computed as

$$\begin{aligned}
\text{FIM}(\xi^*,\boldsymbol{\theta}) &= \text{FIM}(\xi,\boldsymbol{\theta})+\text{FIM}_s^+(\xi,\boldsymbol{\theta})-\text{FIM}_s^-(\xi,\boldsymbol{\theta}),\\
&= \text{FIM}(\xi,\boldsymbol{\theta})+\begin{bmatrix}\text{FIM}_s^{11^+}(\xi,\boldsymbol{\theta})-\text{FIM}_s^{11^-}(\xi,\boldsymbol{\theta}) & \text{FIM}_s^{12^+}(\xi,\boldsymbol{\theta})-\text{FIM}_s^{12^-}(\xi,\boldsymbol{\theta})\\ \text{FIM}_s^{21^+}(\xi,\boldsymbol{\theta})-\text{FIM}_s^{21^-}(\xi,\boldsymbol{\theta}) & \text{FIM}_s^{22^+}(\xi,\boldsymbol{\theta})-\text{FIM}_s^{22^-}(\xi,\boldsymbol{\theta})\end{bmatrix},\\
&= \text{FIM}(\xi,\boldsymbol{\theta})+\begin{bmatrix}\frac{N}{\sigma^2}\mathbf{X}_s^-\mathbf{P_s}(\mathbf{X}_s^+)^T & \frac{N}{\sigma^3}\mathbf{X}_s^-\mathbf{P_s}(\mathbf{U}_s^+)^T\\ (\frac{N}{\sigma^3}\mathbf{X}_s^-\mathbf{P_s}(\mathbf{U}_s^+)^T)^T & \frac{N}{\sigma^4}\mathbf{U}_s^-\mathbf{P_s}(\mathbf{U}_s^+)^T\end{bmatrix},\\
&= \begin{bmatrix}\text{FIM}^{11}(\xi^*,\boldsymbol{\theta}) & \text{FIM}^{12}(\xi^*,\boldsymbol{\theta})\\ \text{FIM}^{21}(\xi^*,\boldsymbol{\theta}) & \text{FIM}^{22}(\xi^*,\boldsymbol{\theta})\end{bmatrix}, \qquad (5.4.3)
\end{aligned}$$

where $\mathbf{X}_s^- = [(\boldsymbol{x}_{1s}^T)^+, -(\boldsymbol{x}_{1s}^T)^-, \boldsymbol{x}_{2s}^T]$ and $\mathbf{X}_s^+ = [(\boldsymbol{x}_{1s}^T)^+, (\boldsymbol{x}_{1s}^T)^-, \boldsymbol{x}_{2s}^T]$ are a $k \times 3$ design matrix corresponding to adding and deleting one alternative at a time, $\mathbf{U}_s^- =$

$[g(\boldsymbol{x}_{1s}^{+}), -g(\boldsymbol{x}_{1s}^{-}), g(\boldsymbol{x}_{2s})]$ and $\mathbf{U}_s^{+} = [g(\boldsymbol{x}_{1s}^{+}), g(\boldsymbol{x}_{1s}^{-}), g(\boldsymbol{x}_{2s})]$ are a row vector of the mean utility values corresponding to the added and deleted alternatives; and $\mathbf{P_s}$ is defined as

$$\mathbf{P}_s = \begin{bmatrix} a & 0 & -a \\ 0 & b & -b \\ -a & b & a-b \end{bmatrix}, \tag{5.4.4}$$

where $a = P_{1s}^{+} - (P_{1s}^{+})^2$, $b = P_{1s}^{-} - (P_{1s}^{-})^2$, and P_{1s}^{-} and P_{1s}^{+} are the corresponding choice probabilities for the deleted and added alternatives, respectively, which is defined as in Equation (2.5.12) for non-death choice sets.

Hence, the $\mathcal{D}_{\mathbf{S}}$-optimality criterion for the modified design, ξ^*, is computed as

$$\mathcal{D}^{B}_{\mathbf{S},\text{FIM}} \;=\; \int \left|\text{FIM}_{\boldsymbol{\beta}}^{-1}(\xi^*, \boldsymbol{\theta})\right|^{1/k} \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{5.4.5}$$

where,

$$\begin{aligned} \text{FIM}_{\boldsymbol{\beta}}(\xi^*, \boldsymbol{\theta}) &= \text{FIM}^{11}(\xi^*, \boldsymbol{\theta}) - \text{FIM}^{12}(\xi^*, \boldsymbol{\theta}) \left[\text{FIM}^{22}(\xi^*, \boldsymbol{\theta})\right]^{-1} \text{FIM}^{21}(\xi^*, \boldsymbol{\theta}), \\ &= \left(\text{FIM}^{11}(\xi, \boldsymbol{\theta}) + \frac{N}{\sigma^2}\mathbf{X}_s^{-}\mathbf{P_s}(\mathbf{X}_s^{+})^T\right) - \left(\text{FIM}^{12}(\xi, \boldsymbol{\theta}) + \frac{N}{\sigma^3}\mathbf{X}_s^{-}\mathbf{P_s}(\mathbf{U}_s^{+})^T\right) \times \\ &\quad \left(\text{FIM}^{22}(\xi, \boldsymbol{\theta}) + \frac{N}{\sigma^4}\mathbf{U}_s^{-}\mathbf{P_s}(\mathbf{U}_s^{+})^T\right)^{-1} \left(\text{FIM}^{12}(\xi, \boldsymbol{\theta}) + \frac{N}{\sigma^3}\mathbf{X}_s^{-}\mathbf{P_s}(\mathbf{U}_s^{+})^T\right)^{T}. \end{aligned}$$

Kessels et al. (2009) suggest that it is faster to compute the Bayesian design criterion value using the Cholesky decomposition of the FIM, which for a positive definite matrix is defined as

$$\text{FIM}(\xi, \boldsymbol{\theta}) = \mathbf{L}^T\mathbf{L}, \tag{5.4.6}$$

where $\mathbf{L}$ is an upper triangular matrix called the Cholesky factor. This is because using the triangular matrix reduces the number of operations required to compute the inverse, or any other functions of the FIM, from inverting the original FIM (since in general inverting a $k \times k$ matrix takes about k^3 operations, whereas using the triangular matrix requires $\frac{k^3}{3}$). This decomposition also makes it much easier to compute the determinant

of the FIM, which is equal to the square of the product of the diagonal elements of the Cholesky Factor $\mathbf{L}$. Thus, the $\mathcal{D}_{\mathbf{S}}$-optimality criterion can be economically calculated using the Cholesky decomposition as

$$\begin{aligned}
\mathcal{D}_{\mathbf{S},\text{FIM}} &= \left|\text{FIM}_{\boldsymbol{\beta}}^{-1}(\xi,\boldsymbol{\theta})\right|^{1/k}, \\
&= \frac{1}{\left|\text{FIM}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta})\right|^{1/k}}, \\
&= \frac{1}{(|\mathbf{L}^T||\mathbf{L}|)^{1/k}} = \frac{1}{\left(\prod_{i=1}^{k} l_{ii}\right)^{2/k}},
\end{aligned} \tag{5.4.7}$$

where l_{ii} is the diagonal element of the Cholesky factor $\mathbf{L}$ of $\text{FIM}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta})$.

The computational time saved by using the Cholesky decomposition to compute the criterion value of a random design using a single draw from the prior distribution of the parameters appear to be negligible, approximately $4e^{-04}$ seconds. However, since we consider using the Bayesian approach, where the criterion values is average over a large numbers of random draws, here 1000 draws, as well as using more than one starting design to find the optimal Bayesian design, then it might be worth using Cholesky decomposition to compute the criterion value and finding the optimal design.

Therefore, it might be more cost effective to compute the Bayesian design criterion value of the modified design based on updating the Cholesky factor for the FIM of the starting design, $\mathbf{L}$, rather than directly solving the inverse in Equation (5.4.5). This factor is updated using the additional matrix to the original FIM matrix defined in Equations (5.4.3), to obtain the corresponding Cholesky factor of $\text{FIM}_{\boldsymbol{\beta}}(\xi^*,\boldsymbol{\theta})$.

The $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ criterion value for the modified design is then computed as follows.

1. For each LHS draw from the given prior distribution for the unknown parameters, $\boldsymbol{\theta}^r = (\boldsymbol{\beta}^r, \sigma^r)$, compute the $\text{FIM}(\xi,\boldsymbol{\theta})$ of the starting design, denoted as $\text{FIM}_r(\xi,\boldsymbol{\theta})$;

2. Update the $\text{FIM}_r(\xi,\boldsymbol{\theta})$ after each single change using the formal in Equation

(5.4.5), to obtain the $\text{FIM}_r(\xi^*, \boldsymbol{\theta})$, and then compute the corresponding Choelsky factor of $\text{FIM}_{\boldsymbol{\beta}}(\xi^*, \boldsymbol{\theta}^r)$, denoted as $\mathbf{L}_r^*$;

3. Compute the criterion value for each draw as

$$\begin{aligned} \mathcal{D}^r_{\mathbf{S},\text{FIM}} &= \left|\text{FIM}^{-1}_{\boldsymbol{\beta}}(\xi^*, \boldsymbol{\theta}^r)\right|, \\ &= \frac{1}{\left(\prod_{i=1}^k l^*_{ii}\right)^{2/k}}, \end{aligned}$$

where l^*_{ii} is the diagonal element of the Choelsky factor $\mathbf{L}_r^*$ of $\text{FIM}_{\boldsymbol{\beta}}(\xi^*, \boldsymbol{\theta})$;

4. Compute the $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ as

$$\hat{\mathcal{D}}^B_{\mathbf{S},\text{FIM}} = \frac{1}{R}\sum_{r=1}^{R} \mathcal{D}^r_{\mathbf{S},\text{FIM}}.$$

We do not need to compute the FIM of the modified design, ξ^*, for each alternative change: instead we update the FIM and the Cholesky factor for the starting design computed for each draw using the additional matrix to the original FIM matrix of the starting design. A similar procedure is used to economically compute the Bayesian $\mathcal{D}_{\mathbf{S}}$ criterion value based on the GFIM for each profile change made by the coordinate-exchange algorithm, where

$$\text{GFIM}(\xi^*, \boldsymbol{\theta}) = \text{GFIM}(\xi, \boldsymbol{\theta}) + \begin{bmatrix} \frac{N}{\sigma^2}\mathbf{X}_s^- \mathbf{P}_{\mathbf{s}} (\mathbf{X}_s^+)^T & \frac{N}{\sigma^3}\mathbf{X}_s^- \mathbf{P}_{\mathbf{s}} (\mathbf{U}_s^+)^T \\ (\frac{N}{\sigma^3}\mathbf{X}_s^- \mathbf{P}_{\mathbf{s}} (\mathbf{U}_s^+)^T)^T & \frac{N}{\sigma^4}\mathbf{U}_s^- \mathbf{P}_{\mathbf{s}} (\mathbf{U}_s^+)^T \end{bmatrix} \tag{5.4.8}$$

and $\mathbf{X}_s^-$, $\mathbf{P}_s$, $(\mathbf{X}_s^+)$, $\mathbf{U}_s^-$ and $\mathbf{U}_s^+$ are as defined in equation 5.4.3.

5.5 Simplifying the Choice Task

The design algorithm described previously produces Bayesian optimal pairwise choice experiments with alternatives that can vary in all attributes under study. This is known as a full profile design, and the design as a Bayesian optimal full profile design. For a large design problem, however, many studies found that increasing the number of the varied attributes within alternatives in a choice task to more than four attributes affects the ability to choose, and hence contributes to an increase in the error variance (e.g., Green, 1974; Arentze et al., 2003; Caussade et al., 2005; and Schwabe et al., 2003).

In addition, respondents' choices are assumed to be made by making a compensatory decision. That is, respondents trade off between attributes and attribute levels such that the unattractive level of an attribute can be compensated by attractive levels of another attribute. However, it has been shown that for choice tasks with large numbers of attributes (more than four), respondents violate the compensatory assumption, since they usually make their choices based on trading off between the level of one or a small subset of attributes while ignoring the others (Kessels et al., 2011a). Thus, the decision making process is dominated by these attributes, which affects the accuracy of the estimated preference values.

To accurately measure respondents' choices, and consequently the preference parameter values, it is more reasonable to simplify the choice task by fixing the level of some attributes in each choice set. For instance, in the pairwise comparison of the AQL-5D health states 11231 and 11423, respondents can ignore the fixed attributes in this choice task and make their choices based on the remaining non-fixed attributes

without violating the compensatory decision assumes main effects. A profile with some fixed attributes is called a partial profile, and the resulting choice design is known as a partial profile design.

To generate Bayesian optimal partial profile designs, one needs first to specify the number of fixed attributes in the choice design. In our design problem for valuing AQL-5D health states, we fix two attributes in each choice task while allowing the remaining three attributes to vary. The fixed attributes should not be the same in all choice sets, so that the preference values for all attributes under study can be evaluated and estimated (Kessels et al., 2011a). Therefore, methods are needed to determine the fixed attributes in each choice set, and the levels of fixed and non-fixed attributes in a choice design as discussed in Kessels et al. (2014) and Cuervo et al. (2015). Sections 5.5.1 and 5.5.2 describe the methods used to specify the fixed attributes in each choice set, and select the levels for fixed and non-fixed attributes in our design problem, respectively.

5.5.1 Determining the Fixed Attributes in Each Choice Set

The selection of fixed attributes in each choice task is based on the total number of fixed attributes and choice sets required in the design. In our design algorithm, we attempt to balance the number of times each attribute is held fixed through the choice design, so that each attribute is evaluated an equal number of times. Also, since we are willing to fix more than one attribute, mainly two attributes in our design problem, we also attempt to balance the number of times each attribute is held constant with another attribute, as in the balanced incomplete block design (BIBD) used in Kessels et al. (2014).

To do so, we specify all possible combinations of fixing particular number of attributes based on the number of attributes under study, say Δ, and the required number of constant attributes in each alternatives, denoted as Δ_f. We then spread these combinations through the choice design based on the total number of choice sets, such

that each combination appears an equal number of times in the choice design. Thus, for a design with S choice sets the number of times each combination appears in the choice design is computed as $S/\binom{\Delta}{\Delta_f}$, where $\binom{\Delta}{\Delta_f}$ is the total number of all possible combinations for fixing some attribute Δ_f out of Δ attributes. This procedure ensures that each fixed attribute and combination of constant attributes appear an equal number of times through the choice design, and hence all attributes under study can be evaluated.

For instance, for an AQL-5D design problem that contains 20 pairwise comparisons for 5 attributes where 2 attributes are fixed in each choice set, there are 10 possible combinations of fixing 2 attributes in each choice set: $\{(1,2),(1,3),(1,4),(1,5),(2,3),(2,4),(2,5),(3,4),(3,5),(4,5)\}$. Each attribute can be fixed four times. However, to satisfy the property of attribute balance through the choice design each combination must occur two times in the design, i.e. $20/10 = 2$. If the total number of choice sets is not divisable by the total number of all possible combinations, says 21 pairwise comparisons, then we repeat each combination two times and randomly select one more combination to proceed in the choice design. Thus, the algorithm is flexible enough to produce partial profile design for any number of choice sets and any number of fixed attributes.

5.5.2 Assigning Attribute levels for Fixed and Non-fixed Attributes

The levels for the constant attributes can be chosen randomly for each choice set. This follows since the main effects do not have any effect on the information provided by the pairwise experiments. The selection of non-fixed attributes is firstly based on the random search algorithm described in Section 5.4.1. Using this algorithm the levels for these attributes are selected randomly such that the Bayesian $\mathcal{D}_{\mathbf{S}}$ criterion is optimised. Then, the coordinate-exchange algorithm is applied over the non-fixed attributes only

in each profile to select the best level for the non-fixed attributes in the choice design such that it optimises the design criterion value as described in Section 5.4.2.

This procedure, however, might produce more constant attributes than required in some choice sets, since for a Bayesian design the number of constant attributes in a choice set can be affected by the prior distribution chosen to optimise the underlying design criterion. Kessels et al. (2011b) illustrate that a prior distribution of the preference parameters that has a large variance and prior mean far away from zero is more likely to produce more fixed attributes in the Bayesian choice designs. Therefore, in order to prevent any additional fixed attributes produced by the prior distribution in our choice designs, we further modify the coordinate-exchange algorithm so that all the levels of the non-fixed attributes are restricted to vary within each choice sets.

Having described our design algorithm, in the following sections, we employ this algorithm to generate a Bayesian optimal pairwise choice experiment for valuing AQL-5D health states within the required QALY scale.

5.6 Bayesian Optimal Choice Design for the AQL-5D Classification System

In this section, we apply our design algorithm to construct Bayesian pairwise optimal choice experiments for the AQL-5D classification system based on the asymptotic Bayesian criteria $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ and $\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$. First, we construct the Bayesian designs using a sensible prior for the preference parameter, $\boldsymbol{\beta}$, that exhibits the expected features of the parameter values such as signs and effect sizes. Since the preference parameters,

$\boldsymbol{\beta}$, are associated with the incremental decrease in the mean utility when moving one level in one attribute of the AQL-5D system, the parameter values are expected to be positive and small in magnitude, as discussed in Section 4.5.1.

For the construction of the Bayesian designs in this chapter, we mainly consider using one type of prior distribution, in particular the independent Gamma(1,10) prior distribution, for each of the preference parameters, β_i for $i = 1, \ldots, 20$, and the scale parameter of the random error, σ. Assuming the same prior distribution for the preference parameters means that permuting the attribute levels of a health state will not change the utilities of the resulting states. For instance, both AQL-5D health states 02314 and 43210 will have the same utility values under this assumption. Nevertheless, this might not be the case in reality as usually a respondent has more preference for one attribute over the other. Therefore, in Chapter 6, a more realistic prior distribution derived from a real AQL-5D data is considered to generate a Bayesian optimal design for the AQL-5D case study, to study the effect of the prior distribution on the choice of Bayesian optimal design.

In this section, we mainly describe the construction of different Bayesian $\mathcal{D}_{\mathbf{S}}$-optimal choice designs with full and partial profiles for the AQL-5D case study using our design algorithm, as will be described in Section 5.6.1. We then compare Bayesian partial profile design to Bayesian full profile design, and our Bayesian designs to the LBD, based on their optimality criterion values and the design efficiencies, as illustrated in Section 5.6.2 and Section 5.6.3, respectively.

5.6.1 Constructing Bayesian Pairwise Experiments for the AQL-5D Classification System

Having specified the prior distribution, we generate different $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ and $\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$ optimal designs with full and partial profiles using our Bayesian design algorithms described in Sections 5.4 and 5.5. In constructing these designs, we restrict the number

of pairwise comparisons to be equal to that in the LBD described in Section 4.2.2, i.e. 32 pairwise comparisons, in order to be able to compare the designs. Also, all Bayesian optimal designs are generated in the same way, so we can compare the resulting designs based on their optimality criterion value. Each Bayesian optimal design is generated as follows:

- The best random Bayesian $\mathcal{D}_{\mathbf{S}}$-optimum design is selected from 1,000 random Bayesian designs with either full or partial profiles (two fixed attributes) using the random search algorithm described in Section 5.4.1. The algorithm returns the choice design that minimises the underlying asymptotic Bayesian criterion value, i.e. either $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ or $\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$ defined in equations 5.2.20 and 5.2.21 for the logit model, respectively, as the best random Bayesian $\mathcal{D}_{\mathbf{S}}$-optimum design. The Bayesian criterion values are computed by averaging the criterion value over the underlying prior distribution, here the Gamma(1,10) prior for each parameter.

- The Bayesian criterion value is estimated using a Monte Carlo method based on a reasonable number of LHS draws from the prior distribution that assures the convergence of the criterion value as will be shown later on this section, where, for R LHS draws, the Bayesian criterion value is estimated as

$$\hat{\mathcal{D}}^B_{\mathbf{S},\text{FIM}} = \frac{1}{R}\sum_{r=1}^{R}\left|\text{FIM}^{-1}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta}^r)\right|^{1/k},$$

  and

$$\hat{\mathcal{D}}^B_{\mathbf{S},\text{GFIM}} = \frac{1}{R}\sum_{r=1}^{R}\left|\text{GFIM}^{-1}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta}^r)\right|^{1/k},$$

  where $\text{FIM}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta})$, and $\text{GFIM}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta})$ are defined as in Equations (5.2.19), (5.2.22). However, before generating and comparing random Bayesian designs, we check the convergence of the estimated design criterion value, $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$, to the true value, $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$. To do so, we investigate the convergence of the $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ criterion value

for a random design generated by our design algorithm using different numbers of draws and sampling methods from the given prior distribution. Figure 5.1 shows the $\hat{\mathcal{D}}^{B}_{\mathbf{S},\mathrm{FIM}}$ criterion values for different numbers of draws from the prior distribution for the unknown logit model parameters, Gamma(1,10), and for the LHS and SMC sampling methods.

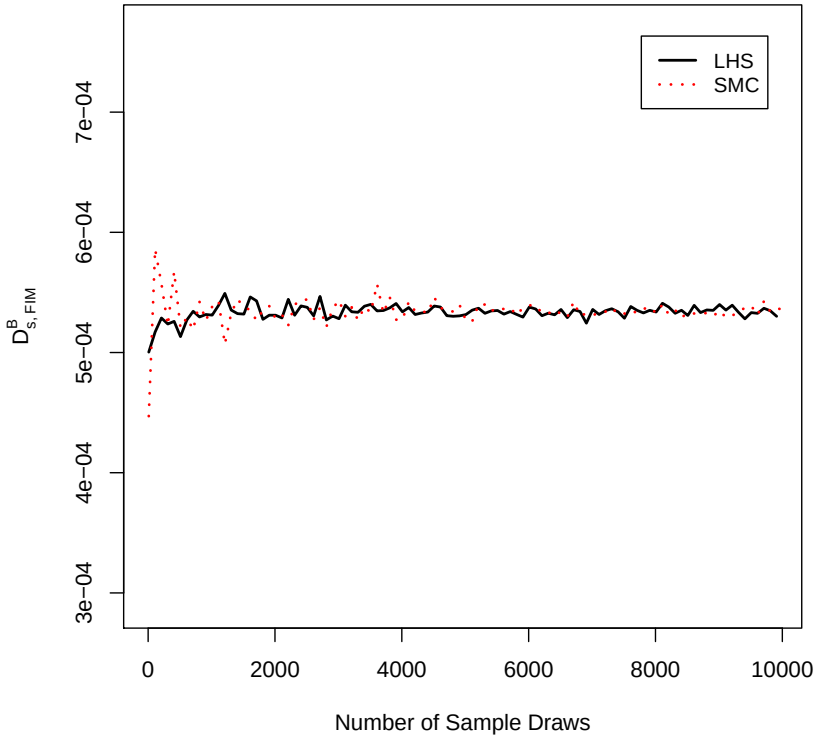

Figure 5.1: Convergence of $\mathcal{D}^{B}_{\mathbf{S},\mathrm{FIM}}$ value using LHS and SMC methods

The plot indicates that the estimated design criterion value converges faster to the 'true' value estimated for 10,000 LHS draws, $5.30e^{-04}$, when using the LHS sample method than it does when the SMC method. For our design problem, it seems that samples of 1,000 LHS draws are enough to approximate the asymptotic Bayesian $\mathcal{D}_{\mathbf{S}}$ criterion value. Therefore, 1,000 LHS samples are used to estimate all the criterion values for the generated Bayesian designs throughout this thesis.

- The coordinate-exchange algorithm is then applied to the best random Bayesian designs with full or partial profiles, generated in the first stage. The algorithm searches for the best levels for all attributes in each profile such that the underling optimality criterion is optimised. This is done through cycling the attribute level of each attribute and alternative in the choice design over all possible options, as described in Section 5.4.2, where the underlying Bayesian criterion is evaluated for each change using the updated Cholesky factor for the $\text{FIM}_{\boldsymbol{\beta}}$ or $\text{GFIM}_{\boldsymbol{\beta}}$ described in Section 5.4.3. The change is executed if it improves the criterion value. The cycling procedure is stopped after evaluating all possible exchanges in the pairwise choice design, and returns the design with the alternatives that optimise the criterion value.

  In our illustration example, the coordinate-exchange algorithm shows improvement in the design efficiency compared of the best random Bayesian design by approximately $30\% - 50\%$ for the partial and full profile designs, respectively. The design returned from this stage using 15 different best random designs is named the optimal or near optimal Bayesian choice design, and it is then used to perform the choice experiment and collect the choice data to estimate the preference parameters.

The following section presents the design criterion values for the generated Bayesian optimal designs with full and partial profiles for the AQL-5D classification system, and then compares these designs based on their optimality criterion values.

5.6.2 Comparing Bayesian Optimal Designs

In this section, we compare Bayesian optimal partial profile choice designs to Bayesian full profile choice designs generated for the AQL-5D case study. The designs are based on $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$ and $\mathcal{D}^B_{\mathbf{s},\text{GFIM}}$ assuming Gamma(1,10) prior distributions, each with 32 pairwise comparisons where each includes one death comparison as shown in the Ap-

pendix A.2.2. Table 5.2 illustrates the criterion values for each design and the convergence of the design criterion value using different number of the LHS draws. We also calculate the corresponding value of the design criterion for which the design was not optimised, to investigate the effect of using different design criteria on the choice of the Bayesian designs.

Table 5.2: Bayesian $\mathcal{D}_{\mathbf{S}}$ design criterion values for full and partial profiles $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$- and $\mathcal{D}^B_{\mathbf{s},\text{GFIM}}$-optimal choice designs evaluated under the Gamma(1,10) prior distribution for the underlying design criterion (shaded values) and the criterion for which the design was not optimised (evaluation criterion), together with the criterion values corresponding to different number of LHS draws evaluated under the underlying design criterion

Design	***Design Criterion***	***Number of LHS Draws***			***Evaluation Criterion***	
		1000	5000	10000	$\mathcal{D}^B_{\mathbf{s},\text{FIM}}$	$\mathcal{D}^B_{\mathbf{s},\text{GFIM}}$
Full profile design	$\mathcal{D}^B_{\mathbf{s},\text{FIM}}$	0.000208	0.000208	0.000209	**0.000208**	0.000194
	$\mathcal{D}^B_{\mathbf{s},\text{GFIM}}$	0.000195	0.000194	0.000194	0.000210	**0.000195**
Partial profile design	$\mathcal{D}^B_{\mathbf{s},\text{FIM}}$	0.000302	0.000301	0.000304	**0.000302**	0.000279
	$\mathcal{D}^B_{\mathbf{s},\text{GFIM}}$	0.000278	0.000278	0.000280	0.000305	**0.000278**

The values of the design criterion estimated for each design in Table 5.2 using increasing numbers of LHS random draws illustrates little variation. This suggests that a sample of 1,000 random draws is adequate to achieve convergence the value of the design criterion to the true value.

Comparing design criterion values for $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$-optimum designs and $\mathcal{D}^B_{\mathbf{s},\text{GFIM}}$ optimum designs with full and partial profiles (shaded values), it can be seen that, for both full and partial profile designs, $\mathcal{D}^B_{\text{GFIM}}$-optimum choice designs have slightly better criterion values. This is because the aim here is to minimise the determinant of the posterior VCM of the preference parameter, $\boldsymbol{\beta}$, and the optimality criterion values of the $\mathcal{D}^B_{\mathbf{s},\text{GFIM}}$-optimum choice designs with full and partial profiles, i.e. 0.000195 and 0.000278, respectively, are smaller compared to the one observed for the $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$-optimum choice designs, i.e. 0.000208 and 0.000302. Nevertheless, to be able to investigate the effect of using different criteria on the choice of the optimal design itself, we evaluate each design under other design criteria for which it was not optimised, as

shown in Table 5.2, and then compare how well each design performs with respect to other criterion.

In our example, Bayesian optimal choice designs optimised for particular design criterion perform as efficient as the optimal choice designs that were not optimised with respect to the underlying design criterion. This is because, for both full and partial profile choice designs, the efficiency loss of Bayesian designs optimised for particular design criterion relative to other choice designs is small, approximately 1%, where the efficiency of a design ξ to the optimal design ξ^* is calculated as in Equation (3.4.6). For Bayesian $\mathcal{D}_{\mathbf{S}}$ design the design efficiency is defined as

$$\mathcal{D}^B_{\mathbf{S},eff} = \frac{\mathcal{D}^B_{\mathbf{S}}(\xi^*, \boldsymbol{\theta})}{\mathcal{D}^B_{\mathbf{S}}(\xi, \boldsymbol{\theta})}. \tag{5.6.1}$$

Bayesian partial profile designs produce a negligible amount of efficiency loss, around 0.9%, when the design is optimised using different design criterion than the evaluation criterion. For instance, the $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ criterion values for Bayesian partial profile design optimised using $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ and $\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$ are similar, 0.000302 and 0.000305, respectively. Therefore, either criteria can be used to construct an efficient Bayesian choice design with full or partial profiles, as shown by our illustration example for the AQL-5D classification system.

To investigate the effect of fixing two attributes in each choice set on the design efficiency, we compare Bayesian partial profile design based on particular design criterion to the corresponding full profile designs. The designs are compared based on their optimality criterion values, using the Bayesian $\mathcal{D}_{\mathbf{S}}$ efficiency measure defined earlier, where here the Bayesian full profile design based on the underlying asymptotic Bayesian criterion represents the optimal choice design, ξ^*. Based on this comparison we could learn how much we lose in design efficiency when the Bayesian partial profile design is used instead of the Bayesian full profile design.

Table 5.3 shows the criterion values for each design, together with their $\mathcal{D}^B_{\mathbf{S},eff}$ val-

Table 5.3: $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ and $\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$ criterion values for Bayesian design with full and partial profiles, and the design efficiency with respect to Bayesian full profile design

Design	***Design Criterion***	
	$\mathcal{D}^B_{\mathbf{S},\text{FIM}}$	$\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$
Full profile design	0.000208	0.000195
Partial profile design	0.000302	0.000278
$\mathcal{D}^B_{\mathbf{S},eff}$	68.87%	70.14%

ues. In general, fixing two attributes in each choice set reduces the design efficiency by approximately $30\% - 32\%$, as shown in Table 5.3, for partial profile designs generated based on $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ and $\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$. Therefore, partial profile designs provide less information on the parameter values compared to Bayesian full profile designs. Nevertheless, one could use $30\% - 32\%$ more respondents in order to obtain the same amount of information and efficiency as in Bayesian full profile designs.

On the plus side, this efficiency loss does simplify the choice task for respondents and hence increase the response efficiency, i.e. decreasing the error in respondents' choices associated with complexity of the choice tasks or any other unobserved cognitive factors that could affect respondents' choices. Thus, this simplification of the choice task gives more accurate measurement for respondents choices which reduces the error variance associated with these measurements, and hence return more reliable estimates for the preference parameters, $\boldsymbol{\beta}$. This might improve the overall design efficiency, and hence the overall precision of the estimated parameter of interested, since the overall design efficiency depends on balancing both the statistical efficiency, i.e. associated with minimising the variance of the parameter estimates, and the response efficiency (Johnson et al., 2013). The statistical efficiency can be improved by asking respondents many difficult choice questions, nevertheless, this might affect the response efficiency and increase the error variance of respondents choices and hence reduce the overall design efficiency. Therefore, in practice, there might be trade-off between maximising the statistical and response efficiencies to obtained best design practises.

Additionally, simplifying the choice questions has the benefit of forcing respondents

to make trade-off between the non-fixed attributes even if a dominant attribute exists (Kessels et al., 2011a). Therefore, the reduction on the design efficiency of $30\% - 32\%$ can be considered as a moderate loss when taking into account the impact of using Bayesian full profile designs in respondents choices and the validity of the compensatory decision assumption of the discrete choice model. This makes the use of Bayesian partial profile optimal design, generated for our the AQL-5D system or even more complex choice design problems, more attractive in reality.

5.6.3 Comparing LBD and Bayesian Optimal Designs

In this section, we compare Bayesian optimal designs based on the $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$ and $\mathcal{D}^B_{\mathbf{s},\text{GFIM}}$ as described in the previous section to the level balanced design (LBD) presented in Section 4.2.2. Assuming that the specified prior distributions for the parameter values, Gamma(1,10) prior, are correct, then we can compare our Bayesian optimal designs with the LBD using the corresponding values of the underlying asymptotic Bayesian criterion for the LBD – either $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$ or $\mathcal{D}^B_{\mathbf{s},\text{GFIM}}$ evaluated at the given prior distribution for the preference parameters. Also, we compare the efficiency of the LBD to the generated Bayesian designs, again using the $\mathcal{D}^B_{\mathbf{s},eff}$ measure defined earlier, to investigate the effect of using LBD instead of Bayesian design optimised for particular design criterion under the suggested prior (Table 5.4).

Table 5.4: $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$ and $\mathcal{D}^B_{\mathbf{s},\text{GFIM}}$ criterion values for the LBD and Bayesian optimal designs with full and partial profiles together with the efficiency of the LBD with respect to the optimal Bayesian full profile design and partial profile design (value between brackets)

Design	***Design Criterion***	
	$\mathcal{D}^B_{\mathbf{s},\text{FIM}}$	$\mathcal{D}^B_{\mathbf{s},\text{GFIM}}$
LBD	0.000685	0.000593
Full profile design	0.000208	0.000195
Partial profile design	0.000302	0.000278
$\mathcal{D}^B_{\mathbf{s},eff}$	30.36% (44.08%)	32.84% (47.04%)

In general, our Bayesian choice designs with full and partial profiles outperform the LBD design. Comparing the LBD with Bayesian full and partial profile designs based on the $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ criterion (Table 5.4), and assuming the specified prior is correct, the indications are that using the LBD design instead of Bayesian optimal designs produces much less information about the preference parameter values. This is because the LBD produces an efficiency loss of 65% and 70% compared to Bayesian partial and full profiles designs, respectively. Therefore, we need more than twice the number of respondents in order to obtain the same amount of information as from the Bayesian optimal design constructed based on $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ and Gamma(1,10) prior. The same kinds of results are obtained when comparing LBD to Bayesian full and partial profile designs based on the $\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$ design criterion, with slightly lower efficiency loss of the LBD design compared to the Bayesian optimal full and partial profile designs.

The outperformance of the LBD by Bayesian choice designs might be related to the method used to construct the choice design as well as the incorporation of the prior information in constructing the choice design. Our design algorithm uses an advanced choice algorithm that allows the incorporation of any suitable prior information in optimising the choice design, and prevents the choice design from having implausible and dominant profiles that might reduce its efficiency. This is in contrast to the LBD that is constructed using the statistical package SAS developed by Huber and Zwerina (1996), which is restricted to orthogonality, level balance and minimal overlap properties, and ignores any prior information about the preference parameters.

Restricting the LBD to these statistical properties, particularly orthogonality, produces three dominant choice tasks and other implausible health states such as AQL-5D health states 00041 and 41000. Respondents might have difficulty in evaluating the illogical alternatives and this could lead to increased error variance, and hence a reduction in the efficiency of the LBD choice design. Also, dominant choice tasks provide no valuable information about the preference parameters as all respondents will choose the dominant health state. Therefore, increasing the number of the dominant choice

Table 5.5: $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ and $\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$ criterion values for LBD and Bayesian optimal designs evaluated at zero prior for the preference parameters, $\boldsymbol{\beta} = \boldsymbol{\beta_0}$

Design	***Design Criterion***	***Evaluation Criterion*** $\mathcal{D}^0_{\mathbf{S}}$
LBD	-	0.000226
Full profile design	$\mathcal{D}^B_{\mathbf{S},\text{FIM}}$	0.000194
	$\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$	0.000194
Partial profile design	$\mathcal{D}^B_{\mathbf{S},\text{FIM}}$	0.000320
	$\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$	0.000324

tasks in a choice design contributes to reducing the design efficiency. Additionally, the orthogonal array used to generate the LBD is based on linear design principles which ignore the nonlinear nature of the choice model in constructing the choice design, and this too affects its efficiency.

The outperformance of LBD by Bayesian choice designs might also depend on the specification of the prior distribution used to construct the Bayesian optimal design. The LBD can be considered as the best utility-neutral design that optimises the $\mathcal{D}^0_{\mathbf{S}}$-optimality criterion, since it satisfies all the three statistical properties; orthogonality, level balance and minimal overlap, as well as the utility balance property when assuming zero prior values for all the preference parameters, $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. Therefore, we compare the efficiency of Bayesian full and partial profile designs generated based on $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ and $\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$ and Gamma(1,10) prior to the LBD assuming that the true parameter values are zero. This can be done by evaluating the design criterion for the Bayesian designs and the LBD at $\boldsymbol{\beta}_0$ as illustrated in Table 5.5, and then comparing the design efficiency of the Bayesian designs to the LBD design based on these values.

The table shows that assuming the true parameter values are zero, $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ and $\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$-optimum full profile designs still outperform the LBD, where the LBD produce and an efficiency loss of approximately 14% compared to these designs. However, the LBD seems to perform better than $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ and $\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$-optimum designs with two constant attributes, as there is approximately 30% efficiency loss when using the partial profile design instead of the LBD if the true parameter values are zero. Also,

Bayesian partial profile designs do not have dominant choices, whereas the LBD has three dominant choice tasks, see A, that provide no valuable information to aid estimating the preference parameters. In addition, simplifying the choice design allows us to obtain information about all the attributes under study even with the presence of dominant attributes, through holding the dominant attribute constant in some choice sets and then trading off between the non-fixed attributes.

5.7 Summary and Discussion

This chapter illustrates how to construct a Bayesian optimal choice design for a health evaluation study, in particular for the AQL-5D case study. Constructing optimal designs requires identifying the discrete choice model and the corresponding design criterion. The chapter started by deriving the asymptotic Bayesian design criteria $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ and $\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$ for our design model, the logit model including the death state defined in Section 4.4.2, then presented our attempt to construct a Bayesian choice design for the AQL-5D case study using available design software such as SAS, JMP and Ngene. However, we have not managed to generate Bayesian optimal choice designs for valuing AQL-5D health states using these programs, as they are limited in handling our design problem, particularly in terms of including the death state and optimising the correct design criterion.

Therefore, we derived a new method to generate a Bayesian optimal choice design for provision of health state utilities within the QALY scale using the latest advanced Bayesian design algorithms available in the design literature. Our design algorithm consists of two stages: first generating designs using a random search algorithm through a

large number of random designs of an appropriate form, and then improve the best random Bayesian choice design, returned by the first stage, using the coordinate-exchange algorithm of Meyer and Nachtsheim (1995). We modified this algorithm so that it handles our design problem. The exchange algorithm finds the best level for each attribute in each alternative in the design through cycling the level of an attribute over all the possible options, and returns the one that improves the criterion value, i.e. that minimises the $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ or $\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$ criterion value. The time required to employ the exchange algorithm over a choice design increases with the number of attributes and attribute levels under study as well as with the required number of choice sets. To reduce the computation time of the exchange procedure, we used an update formula for the FIM and GFIM matrices to economically compute the Bayesian design criterion value for each exchange.

We simplified the choice task by holding two attributes constant in each choice set. Thus, we modified our design algorithm to account for the fixed attributes in the choice design, as described in Section 5.5. An application for our design algorithm is then given in this chapter by generating Bayesian optimal pairwise comparisons with full and partial profiles for the AQL-5D case study. The Bayesian designs are constructed based on the asymptotic Bayesian criteria $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ and $\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$ and assuming Gamma (1,10) prior distribution for each of the unknown logit model parameters. We compared our designs with an existing level balanced design (LBD), previously used in a discrete choice experiment.

Our illustration study indicates that our Bayesian designs are not sensitive to the design criterion used to optimised the choice design. That is, both $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ and $\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$ optimum designs produce approximately the same amount of information about the preference parameters, with a negligible reduction of $0.4\%-1.0\%$ in the design efficiency when using a criterion other than the one used in the optimisation procedure, as shown in Table 5.2. In addition, simplifying the choice task by fixing two attributes in each choice set reduces the design efficiency by $30\% - 32\%$ compared to Bayesian optimal

design with non-fixed attributes. However, this reduction is not remarkable compared with the detrimental effect that Bayesian full profile designs have on the respondents' choices and consequently the reliability of the estimated preference parameter values.

Our study also shows that, if the true preference parameter values are not zero, then using LBD design instead of $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ a or $\mathcal{D}^B_{\mathbf{S},\text{GFIM}}$ optimum designs with full and partial profiles produces a large amount of design efficiency loss, of more than 60%. This reduction is smaller when the true parameter values are zero; however, our Bayesian optimal designs are still the better design options. This reflects the benefit of using more advanced choice design algorithms to generate the choice design, allowing the incorporation of prior information about the parameter values in constructing the choice design. However, the results in our illustration example might depend on the prior distribution specified for optimising the criterion value, and hence finding the optimal choice design. Therefore, in the following chapter, we further investigate the effect of the prior distribution in the choice of the optimal design, and how better or worse priors might affect the design efficiency.

Chapter 6

Sensitivity of the Bayesian Optimal Choice Design to the Prior Distribution

6.1 Introduction

Optimum designs for nonlinear models, more specifically for logit choice models, depend on the values of unknown model's parameters, $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \sigma\}$. In the design literature, this dependency is overcome by using a prior point estimate, $\boldsymbol{\theta_p}$, or a prior distribution $\pi(\boldsymbol{\theta})$ for the unknown parameter vector $\boldsymbol{\theta}$. This yields locally and Bayesian optimum designs, respectively, as mentioned in Section 3.5. Bayesian choice designs incorporate prior information into the design by taking the expectation of the underlying design criterion over the prior distribution. In this chapter, we investigate the effect of the

prior distribution on the Bayesian optimal design by comparing different Bayesian designs optimised with respect to different prior distributions.

The chapter begins with presenting different methods to select a suitable prior distribution for each parameter of the logit model to generate Bayesian optimal choice designs for health economic evaluation studies, particularly for the AQL-5D case study. In Section 6.3, we discuss the considerations of the efficient choice task to optimise the information obtained from the choice data, and how the prior distribution might affect the selection of efficient design choices. In Section 6.4, we demonstrate the effect of the choice of the prior distribution on the choice of the Bayesian optimal designs by comparing Bayesian designs optimised for different prior distributions. Section 6.5 presents a summary and the main findings of this chapter.

All analysis in this chapter is performed for Bayesian pairwise choice designs with partial profiles constructed for the AQL-5D classification system. These designs are calculated for the logit model based on Bayesian $\mathcal{D}_{\mathbf{S}}$-optimality criterion derived in Section 5.2.

6.2

Prior Distributions for Designing Optimal Choice Experiment for Valuing Health States Utilities

Constructing Bayesian optimal designs for choice models requires specifying prior distributions for the unknown model's parameters (unlike the equivalent process for linear

models). The posterior variance-covariance matrix (VCM) of the parameters of interest, here $\boldsymbol{\beta}$, derived based on the Fisher information matrix (FIM) or the generalised Fisher information matrix (GFIM), depends on the unknown parameter values through the choice probability and the mean utility values of the alternatives in the choice task, hence so does the optimality criterion and the generating of the choice designs, as illustrated for the logit model in Section 5.2. So, Bayesian choice designs are computed by averaging the underlying design criterion over a given prior distribution as follows:

$$\begin{aligned}
\mathcal{D}^B_{\mathbf{S},\text{FIM}} &= \int \left|\text{FIM}^{-1}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta})\right| \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}, \\
&= \int \left| \frac{1}{N} \left\{ \sum_{s=1}^{S} \text{FIM}^{11}_s(\xi,\boldsymbol{\theta}) - \sum_{s=1}^{S} \text{FIM}^{12}_s(\xi,\boldsymbol{\theta}) \times \right.\right. \\
&\quad \left.\left. \left[\sum_{s=1}^{S} \text{FIM}^{22}_s(\xi,\boldsymbol{\theta})\right]^{-1} \left[\sum_{s=1}^{S} \text{FIM}^{12}_s(\xi,\boldsymbol{\theta})\right]^T \right\}^{-1} \right|^{1/k} \pi(\boldsymbol{\theta}) d\boldsymbol{\theta},
\end{aligned}$$

and for the GFIM as

$$\begin{aligned}
\mathcal{D}^B_{\mathbf{S},\text{GFIM}} &= \int \left|\text{GFIM}^{-1}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta})\right| \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}, \\
&= \int \left| \frac{1}{N} \left\{ \sum_{s=1}^{S} \text{GFIM}^{11}_s(\xi,\boldsymbol{\theta}) - \sum_{s=1}^{S} \text{GFIM}^{12}_s(\xi,\boldsymbol{\theta}) \times \right.\right. \\
&\quad \left.\left. \left[\sum_{s=1}^{S} \text{GFIM}^{22}_s(\xi,\boldsymbol{\theta})\right]^{-1} \left[\sum_{s=1}^{S} \text{GFIM}^{12}_s(\xi,\boldsymbol{\theta})\right]^T \right\}^{-1} \right|^{1/k} \pi(\boldsymbol{\theta}) d\boldsymbol{\theta},
\end{aligned}$$

where

$$\text{FIM}^{11}_s(\xi,\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^2}(\boldsymbol{x}_{1s} - \boldsymbol{x}_{2s})^T P_{1s}(1-P_{1s})(\boldsymbol{x}_{1s} - \boldsymbol{x}_{2s}), & \text{if } \boldsymbol{x}_d \notin C_s; \\ \\ \frac{1}{\sigma^2} \frac{P_{Js}}{(1-P_{Js})} \left[\exp\left(\frac{g(\boldsymbol{x}_{1s})+\mu}{\sigma}\right)\right]^2 \boldsymbol{x}^T_{1s}\boldsymbol{x}_{1s}, & \text{if } \boldsymbol{x}_d \in C_s, \end{cases}$$

$$\text{FIM}_s^{12}(\xi,\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^3}(\boldsymbol{x}_{1s}-\boldsymbol{x}_{2s})^T P_{1s}(1-P_{1s})\left[g(\boldsymbol{x}_{1s})-g(\boldsymbol{x}_{2s})\right], & \text{if } \boldsymbol{x}_d \notin C_s; \\ \frac{1}{\sigma^3}\left\{P_{1s}^2(1-P_{1s})g(\boldsymbol{x}_{1s}) - \sigma P_{1s}(1-P_{1s}) + \right. & \\ \left. \frac{P_{Js}}{1-P_{Js}}(g(\boldsymbol{x}_{1s})+\mu)\left[\exp\left(\frac{g(\boldsymbol{x}_{1s})+\mu}{\sigma}\right)\right]^2\right\}\boldsymbol{x}_{1s}^T, & \text{if } \boldsymbol{x}_d \in C_s, \end{cases}$$

and

$$\text{FIM}_s^{22}(\xi,\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^4}P_{1s}(1-P_{1s})\left[g(\boldsymbol{x}_{1s})-g(\boldsymbol{x}_{2s})\right]^2, & \text{if } \boldsymbol{x}_d \notin C_s; \\ \frac{1}{\sigma^4}\left\{P_{1s}[g(\boldsymbol{x}_{1s})^2 P_{1s} - g(\boldsymbol{x}_{1s})^3 P_{1s}^3] - 2\sigma P_{1s}(1-P_{1s})g(\boldsymbol{x}_{1s}) + \right. & \\ \left. \frac{P_{JS}}{1-P_{JS}}\left[(g(\boldsymbol{x}_{1s})+\mu)\exp\left(\frac{g(\boldsymbol{x}_{1s})+\mu}{\sigma}\right)\right]^2\right\}, & \text{if } \boldsymbol{x}_d \in C_s, \end{cases}$$

and the sub-matrices of the partitioned GFIM can be obtained from the partitioned FIM as defined in Equation (5.2.23).

The row vector $(\boldsymbol{x}_{1s}-\boldsymbol{x}_{2s})$ represents the differences between the attribute levels of the first and the second alternatives presented in the choice set $C_s = \{\boldsymbol{x}_{1s}, \boldsymbol{x}_{2s}\}$. P_{1s} and P_{Js} are the logit choice probabilities of the first state in the non-death choice set and death state in the death comparison, respectively, which can be deduced from the MNL choice probability defined in Section 2.5.3 as

$$P_{1s} = \frac{1}{1+\exp\left(\frac{g(\boldsymbol{x}_{2s})-g(\boldsymbol{x}_{1s})}{\sigma}\right)},$$

and

$$P_{Js} = \exp\left[-\exp\left(\frac{g(\boldsymbol{x}_{1s})+\mu}{\sigma}\right)\right],$$

where $g(\boldsymbol{x}_{1s}) = 1 - \boldsymbol{\beta}\boldsymbol{x}_{1s}^T$ is the population mean utility for health state $\boldsymbol{x}_{1s}$, defined as a linear function of the logit model parameters, $\boldsymbol{\beta}$. The parameter σ is the scale parameter of the random component of the utility, ϵ.

In Section 3.2.2, we showed that the majority of the DCEs conducted in health economic evaluation study to value health state utilities are constructed based on orthogonal array designs or optimal design principles, ignoring the dependency of the

choice designs on the parameter values. This is usually done by assuming a zero prior for the preference parameters, i.e. replacing the preference parameters of the choice model $\boldsymbol{\beta}$ with a zero vector $\boldsymbol{\beta_0}$. However, this assumption is unrealistic, since it is assumed that respondents have no preference for attribute levels across all choices, i.e. all presented options have equal choice probability. In Chapter 5, we illustrated that incorporating the prior information in generating the choice design improves the efficiency of the design, where a Bayesian choice design is observed to outperform the level balanced design (LBD) that ignores any prior information about the preference parameters. Noticeably, the performance of the choice design depends in some way on the prior distribution as it has to be incorporated when generating the choice tasks such that the design criterion is optimised. Therefore, in this chapter, we investigate the effect of the choice of different prior distribution on the choice of Bayesian optimal design.

To perform this investigation, we select different prior distributions to construct different Bayesian choice designs, particularly for the AQL-5D cases study. The priors are chosen by recalling the methods used to assign a suitable prior distribution for the unknown parameters of the logit model, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)$, in Section 4.5.1 and the Bayesian analysis of the TTO data presented in Section 4.5.3. Prior distributions are assigned then based on these methods:

1. simple prior judgment, based on understanding of the AQL-5D;
2. prior from a previous study on the same health state instrument, i.e. the AQL-5D.

The first method is based on summarising existing knowledge of the unknown parameters, such as the sign and the range of the parameter values, into a prior distribution. In Section 4.5.1, we argued that the value of the preference parameters, $\boldsymbol{\beta}$, are not likely to exceed one and should be positive, since they are associated with the incremental loss in the mean utility when moving one level on one attribute of the

AQL-5D. Also, a change in a level of one attribute is not expected to produce a change in the mean utility greater than the change from perfect health to death, which is one. Similarly, the scale parameter σ is also almost certainly less than one, since the utilities obtained from different individuals for the same health state are not expected to deviate by a large amount (e.g. more than 1). In this chapter, we consider different simple prior distributions that range from 0 to 1. Following Cain (2011), we consider a set of independent Gamma and Uniform prior distributions for each element of the parameter vector $\boldsymbol{\theta}$ as illustrated in Section 4.5.1. In particular, we consider Gamma (1,10), Gamma(5,15) and Beta(1,10) prior distributions, as well as the U[0,1] prior distribution to illustrate the effect of the extreme case where the prior information about individuals' preferences for attribute levels, $\boldsymbol{\beta}$, contains a substantial amount of uncertainty. In addition, other uniform prior distributions with smaller and larger prior mean values of the parameter are considered, to investigate the effect of the range of the parameter values on the choice of Bayesian designs.

The second method uses a Bayesian approach where the prior distribution of the unknown model's parameters, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)$, is derived by analysing a real health care data set from a previous experiment for the same health instrument of interest, the AQL-5D, in a Bayesian manner. The posterior distribution of each parameter can then be used as a prior distribution for that parameter when generating the Bayesian optimal choice designs. In this section, we consider the Bayesian analysis illustrated in Section 4.5.3 for the TTO preference data collected for different AQL-5D health states to assign more realistic prior distributions for the preference and scale parameters. The posterior distributions of the model's parameters are obtained using a Markov Chain Monte Carlo (MCMC) sampling method using different prior distributions such as the Gamma(1,10), Gamma(5,15) and U[0,1] (see Table 4.6 in Section 4.5.3 for a summary of the mean and the 95% posterior interval of the model parameters obtained under each prior distribution). In this section, we use 1,000 draws from the posterior distribution of each parameter generated using Gamma(1,10) prior to set as a prior distribution for that parameter when generating the Bayesian choice design for the AQL-5D case

study. For simplicity, we denote the prior distribution obtained from this analysis as the TTO prior throughout this chapter.

Having specified different prior distributions for the model parameters, it is of interest to investigate the effect of the choice of the prior distribution on the choice of the Bayesian optimal design, i.e. the sensitivity of the Bayesian optimal choice design to the prior distribution used in optimising the underlying Bayesian design criterion value. To understand the effect of the prior distribution on the choice of Bayesian optimal choice design, it is worthwhile considering the role of the prior distribution in selecting efficient choice designs and, hence, optimising the information obtained from the Bayesian choice designs - as illustrated in the following section.

6.3 Appropriate Choice Tasks and the Prior Distribution

In this section, we discuss how the choice of the prior distribution might affect the selection of the choice design questions, and hence the choice of the Bayesian $\mathcal{D}_{\mathbf{S}}$-optimum designs. We first illustrate the main choice task considerations, and then demonstrate the role of the prior distribution in the selection of efficient choice tasks by considering the $\mathcal{D}_{\mathbf{S}}$-optimality design criterion defined in the previous section for the logit choice model.

The $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$-optimality design criterion seeks to minimise the determinant of the

posterior VCM of the preference parameters, $\boldsymbol{\beta}$, that is defined as

$$\begin{aligned}
\mathcal{D}^{B}_{\mathbf{S},\text{FIM}} &= \int \left|\text{FIM}^{-1}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta})\right| \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}, \\
&= \int \left| \frac{1}{N} \left\{ \sum_{s=1}^{S} \text{FIM}^{11}_{s}(\xi,\boldsymbol{\theta}) - \sum_{s=1}^{S} \text{FIM}^{12}_{s}(\xi,\boldsymbol{\theta}) \times \right.\right. \\
&\quad \left.\left. \left[\sum_{s=1}^{S} \text{FIM}^{22}_{s}(\xi,\boldsymbol{\theta})\right]^{-1} \left[\sum_{s=1}^{S} \text{FIM}^{12}_{s}(\xi,\boldsymbol{\theta})\right]^{T} \right\}^{-1} \right|^{1/k} \pi(\boldsymbol{\theta}) d\boldsymbol{\theta},
\end{aligned}$$

This is equivalent to maximising the determinant of the $\text{FIM}_{\boldsymbol{\beta}}(\xi,\boldsymbol{\theta})$ which is obtained by maximising the information regarding the preference parameters, i.e. FIM^{11}_{s}, gain from each choice task, $C_s = \{\boldsymbol{x}_{1s}, \boldsymbol{x}_{2s}\}$, which is define as in Section 6.2.

For death choice sets, more information is obtained from the choice tasks $C_s = \{\boldsymbol{x}_{1s}, \boldsymbol{x}_{d}\}$ as health state $\boldsymbol{x}_{1s}$ becomes more comparable to the state of being dead and respondents will be able to trade between the states. This is possible by comparing death to the worse health state defined by the AQL-5D classification system, as discussed in (Brazier et al., 2009) and as will also be considered in this thesis.

To identify the essential condition on the alternatives of the non-death choice sets, consider the following two cases for the choice set $C_s = \{\boldsymbol{x}_{1s}, \boldsymbol{x}_{2s}\}$.

1. If the alternatives in the choice set are far apart from each other, i.e. $\|\boldsymbol{x}_{1s} - \boldsymbol{x}_{2s}\|$ is large and the alternatives have large differences between their attribute levels, then the choice task will provide no valuable information about the mean utility difference and consequently the preference parameters. This is because such a choice task is considered as one-sided and provides extreme choice probabilities for the alternatives, i.e. $P_{1s} = 1$ or 0, since the logit choice probability is a function of this differences as defined below

$$P_{1s} = \frac{1}{1 + \exp\left(\frac{\boldsymbol{\beta}(\boldsymbol{x}_{1s} - \boldsymbol{x}_{2s})^{T}}{\sigma}\right)}.$$

Thus, a large difference between alternatives results in

$$\begin{aligned}
\text{FIM}_s^{11}(\xi, \boldsymbol{\theta}) &= \frac{1}{\sigma^2}(\boldsymbol{x}_{1s} - \boldsymbol{x}_{2s})^T P_{1s}(1 - P_{1s})(\boldsymbol{x}_{1s} - \boldsymbol{x}_{2s}) = 0, \\
\text{FIM}_s^{12}(\xi, \boldsymbol{\theta}) &= \frac{1}{\sigma^3}(\boldsymbol{x}_{1s} - \boldsymbol{x}_{2s})^T P_{1s}(1 - P_{1s}) \left[g(\boldsymbol{x}_{1s}) - g(\boldsymbol{x}_{2s})\right] = 0, \\
\text{FIM}_s^{22}(\xi, \boldsymbol{\theta}) &= \frac{1}{\sigma^4} P_{1s}(1 - P_{1s}) \left[g(\boldsymbol{x}_{1s}) - g(\boldsymbol{x}_{2s})\right]^2 = 0,
\end{aligned}$$

and, hence, minimises the information obtained about $\boldsymbol{\beta}$, i.e. $\text{FIM}_{\boldsymbol{\beta}} \longrightarrow 0$.

2. If the alternatives are closed from each other, i.e. $\|\boldsymbol{x}_{1s} - \boldsymbol{x}_{2s}\| \longrightarrow 0$, then this also would provide less information from the choice task. This follows since such choice question means comparing the same health states or states with many overlap (i.e. choices with the same level for most attributes) such as $C_s = \{11220, 11221\}$, which will provide no or less information. Hence $\text{FIM}_{\boldsymbol{\beta}} \longrightarrow 0$ as $\|\boldsymbol{x}_{1s} - \boldsymbol{x}_{2s}\| \longrightarrow 0$.

Both these types of choice questions are considered to be poor and inefficient design choices. To produce more informative choice task it would be better to increase the discrepancy between the two alternatives, as long as it does not lead to extreme choice probability, i.e. P_{1s} is not close to one or zero.

Since optimising the $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ design criterion depends on the prior distribution through the definition of the logit choice probability, then the conditions on the attribute level difference between alternatives, $(\boldsymbol{x}_{1s} - \boldsymbol{x}_{2s})$, to avoid extreme choices, might be affected by the choice of the prior distribution, and consequently does the selection of the efficient choice task that optimise the design criterion values. Thus, if in a choice study some attributes are more important than the others, more efficient choices would be obtained by keeping the most important attributes similar while increasing the difference between the less important attributes.

For instance, consider the two choice questions $C_1 = \{30023, 10024\}$ and $C_2 = \{20021, 10023\}$, if the fifth attribute corresponding to the activities attribute in the AQL-5D classification system is more important than the first attribute, as shown

by the TTO prior distribution for this attribute, then the first choice task will be more appropriate than the second choice set. This is because the second choice set has a larger difference between the levels of the activities attribute but smaller level difference between the less important attribute (i.e. the first attribute), which makes this attribute dominates the choices. The opposite is also true: if the first attribute is more important than the fifth attribute, then the second choice set is more appropriate. This makes intuitive sense in terms of individual's preference perspective, as increasing the level difference between the more important attributes while reducing the level difference between the less important attributes results in one-sided choices - choices with extreme choice probabilities - which reduces the design efficiency, as discussed earlier in this section. Hence, different priors would favour different choice sets.

In the case where the attributes are judged a priori to be equally important, we conjecture that choices with similar level differences between all attributes in the two alternatives will be more efficient. However, it is less clear how the conditions on the attribute level differences will be affected by different choice of identically distributed priors for the preference parameters.

The role of the prior distribution in selecting the appropriate choice task suggests that the selection of the efficient $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$ design choices might depend on the prior distribution chosen for optimising the design criterion, and hence misspecifying the prior distribution might lead to a substantial loss in the design efficiency particularly for non-identical prior distributions of the parameters.

As for the $\mathcal{D}^B_{\mathbf{s},\text{GFIM}}$-optimal designs, we would expect similar effects of the prior on the choice of Bayesian designs. This is because this design criterion seeks to minimise the posterior VCM which is equivalent to maximising the determinant of the $\text{GFIM}_{\boldsymbol{\beta}}(\xi, \boldsymbol{\theta})$ over a suitable prior distribution, and this is itself a function of the submatrices of the partitioned FIM as

$$\text{GFIM}(\xi, \boldsymbol{\theta}) = \text{FIM}(\xi, \boldsymbol{\theta}) + \Sigma^{-1},$$

where FIM is the partitioned Fisher information matrix, and Σ is the prior VCM of the MNL model parameters as defined in Equation (5.2.23). In Section 5.6.2, we found that Bayesian choice designs optimised using both criteria based on the Gamma(1,10) prior distribution perform at approximately the same level of efficiency: hence choices optimised by the $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$ will perform well according to the $\mathcal{D}^B_{\text{GFIM}}$ design criterion. We would expect a similar effect of the prior on the choice of the Bayesian designs. Thus, in the following section we investigate the effect of the choice of the prior distribution on the choice of Bayesian designs considering the $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$ design criterion only.

6.4 Sensitivity to the Prior Distribution: Illustrative Study

This section provides an illustrative study that investigates the effect of the choice of prior distribution on the choice of Bayesian optimal choice design, and to what extent using more sensible priors might improve the choice and efficiency of Bayesian designs. This is illustrated by generating $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$-optimum choice designs for different choices of prior distributions, and examining how Bayesian designs optimised for particular prior distribution perform in terms of other priors.

The study is conducted by computing $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$-optimum pairwise choice designs for the AQL-5D classification system using different forms of simple prior distributions of the unknown parameters of the logit model, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)$. These priors are selected based on the available information about the relative size and sign of the parameters, as

discussed in Section 6.2. Each design consists of 32 pairwise comparisons with partial profiles, two constant attributes, constructed using our design algorithm as described previously in Section 5.6.1. Thus, each choice design is generated by firstly using the random search algorithm to obtain the best Bayesian random choice design out of 1,000 random designs, and then by applying the coordinate-exchange algorithm to select the best attribute levels for each alternative in the best random design such that the $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$ optimality criterion is optimised over the given prior distribution. We then compare the efficiency of the Bayesian choice design optimised for a particular prior distribution, referred to as the 'design' prior, relative to other choice designs that were not optimised for the underlying prior distribution. This is done by comparing the design criterion values evaluated for a particular prior distribution, referred to as the 'evaluation' prior, with respect to different Bayesian choice design, where a smaller value gives a more efficient designs for that prior.

Additionally, to study whether a more sensible prior distribution of the parameters improves the choice and the efficiency of the Bayesian design, we compare the resulting $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$ choice designs optimised with respect to different simple prior distributions to the one optimised based on the TTO prior, i.e. the prior distribution obtained from real TTO data for the AQL-5D system, presented in Chapter 4.

In this section, we first investigate the effect of non-identical prior distributions of the parameters on the choice of the $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$-optimum choice designs (Section 6.4.1), and then consider the case where the parameters are identically distributed a priori (Section 6.4.2). In Section 6.4.3, we study the effect of the choice of the prior distribution of the scale parameter, σ, on the choice of the Bayesian designs, and illustrate how a poorly defined prior for this parameter might lead to less efficient design choices.

6.4.1 Comparing Design Efficiency for Non-identical Priors of the Parameters

In this section, we investigate how $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$ choice designs optimised for i.i.d priors of the parameters perform when the 'true' prior distributions of the parameters are not identical, i.e. when the AQL-5D attributes are not equally important in reality, and vice versa.

For illustration, we generate Bayesian pairwise choice designs for identical and non-identical prior distributions of the logit model parameters, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)$, and then compare the resulting designs based on their optimality criterion evaluated for particular prior distribution. For non-identical prior distributions of the parameters, we assume that the last two attributes in the AQL-5D system, sleep and activities, are more important for individuals than the other attributes. Therefore, we allow the preference parameters associated with these attribute level differences, that is $\beta_{13}, \ldots, \beta_{20}$, to have more weight than those associated with the first three attributes, that is $\beta_1, \ldots, \beta_{12}$. Another non-identical prior distribution is represented by the one obtained from the TTO data where the activities attribute seems to be more important than the remaining attributes. Thus, in our illustrative example, we consider the following prior distributions:

- prior 1: $\beta_1, \ldots, \beta_{20} \overset{iid}{\sim} U[0, 1]$, and $\sigma \sim U[0, 1]$
- prior 2: $\beta_1, \ldots, \beta_{12} \overset{iid}{\sim} U[0, 0.2]$, $\beta_{13}, \ldots, \beta_{20} \overset{iid}{\sim} U[0.9, 1]$ and $\sigma \sim U[0, 1]$.
- prior 3: $\beta_1, \ldots, \beta_{20}$ and σ follow the TTO prior distribution.

Table 6.1 presents the $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$ criterion values of Bayesian choice designs optimised for the proposed prior distributions of the parameters together with the corresponding criterion values for each design evaluated under the other prior distributions that have not been used in the optimisation procedure. Thus, for each design we compute the

Table 6.1: $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ design criterion values for different $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$-optimum choice designs evaluated at the design prior (shaded value) and the other suggested prior distributions together with the efficiency of the optimal design with respect to other designs

$\mathcal{D}^B_{FIM}$ ***Optimal Design***	***Design Prior***	***Evaluation Prior***		
		U[0,1]	$U[0, 0.2]\&U[0.9, 1]$	TTO
I	U[0,1]	**0.0042***	0.0071	$3.44e^{-05}$
II	$U[0, 0.2]\&U[0.9, 1]$	0.0106	**0.0028***	$4.30e^{-05}$
III	TTO	0.0063	0.0082	**2.87**e^{-05}*
$\mathcal{D}^B_{\mathbf{S},eff}$	-	39.62%	39.43%	83.43%
	-	66.66%	34.15%	66.74%

* The best criterion value in each column.

Bayesian criterion value under the three priors using Monte Carlo simulation based 1,000 LHS draws from the suggested prior distribution as

$$\hat{\mathcal{D}}^B_{\mathbf{S},lm} = \frac{1}{1000}\sum_{r=1}^{1000}\left|\text{FIM}^{-1}_{\boldsymbol{\beta}}(\xi_l, \boldsymbol{\theta}^r_m)\right|^{1/k}, \quad \text{for} \quad l, m = 1, 2, 3, \tag{6.4.1}$$

where l and m are indexes refer to the 'design' and the 'evaluation' priors, respectively.

To study the sensitivity of the Bayesian choice design to the choice of the prior distribution, we compare the value of $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ for a particular prior distribution with respect to different Bayesian choice designs. The ratio between any two criterion values for particular prior represents the relative efficiency of one design over another.

Considering the design criterion values of the i.i.d U[0,1] priors corresponding to different choice designs presented in Table 6.1, it can be seen that design II optimised using the non-identical prior distribution, prior 2, is less efficient than design I optimised under the i.i.d U[0,1] with an efficiency loss of around 60%. Thus, using design II instead of design I when the 'true' prior distributions of the parameters are i.i.d U[0,1] requires more than twice the number of respondents used in a study in order to achieve the same expected error around the parameter values as in design I. Similarly, if the 'true' prior distributions of the parameters are not identically distributed

as in prior 2, design I performs badly in terms of this prior. This design reduces the efficiency by more than a half relative to the design optimised directly for prior 2, i.e. design II.

Now comparing the efficiency of Bayesian choice design optimised for the TTO prior distributions, design III, relative to both designs I and II, it is observed that design III is less efficient than both designs with respect to their design priors, i.e. priors 1 and 2. Nevertheless, design III performs worse for the non-identical U[0,1] prior than the i.i.d U[0,1] prior, as it leads to an efficiency loss of 65% compared to 33% for the i.i.d U[0,1] prior distribution. Also, if the 'true' prior of the parameters is the TTO prior distribution, then using design I or design II instead of design III reduces the efficiency of the choice design by 16% and 33%, respectively.

A concern with these results is that the observed effect of the prior distribution on the design criterion and hence the Bayesian choice design might be related to the starting design used in the optimisation procedure rather than the prior distribution itself. To asses the importance of this concern, we perform a robustness analysis by replicating the same analysis provided earlier fifteen times, each with a different starting design used in the optimisation. Each run of the design algorithm with respect to a different starting design is denoted as a 'try', hence we have fifteen tries in total for each prior distribution. For simplicity, we display the design criterion values of each prior distribution with respect to the resulting Bayesian designs from each try in a scatter plot instead of a table, as shown in Figure 6.1 for the set of priors presented earlier in this section.

The plot indicates that, for each prior distribution, the design criterion values with respect to different tries are similar to those shown in Table 6.1 with a little variation in these values. This reflects the fact that the observed effect of the prior distribution on the design criterion values is not really related to the choice of the starting design used in the optimisation procedure, but it rather to the efficient design choices selected based on the type of prior distribution used in optimising the choice design.

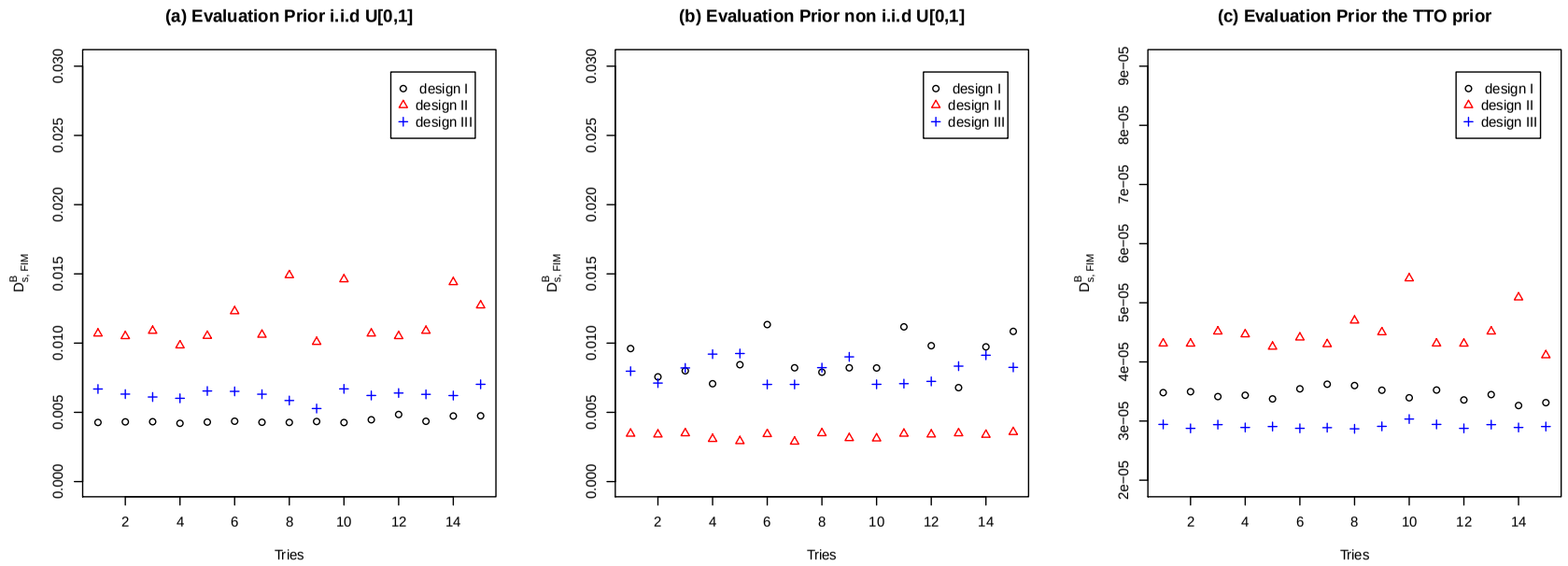

Figure 6.1: Panel plots of the $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$ design criterion values computed for different prior distributions with respect to different Bayesian designs, where designs I, II and III are the corresponding Bayesian choice designs obtained based on the i.i.d U[0,1], the non-identical U[0,1] and the TTO priors, respectively

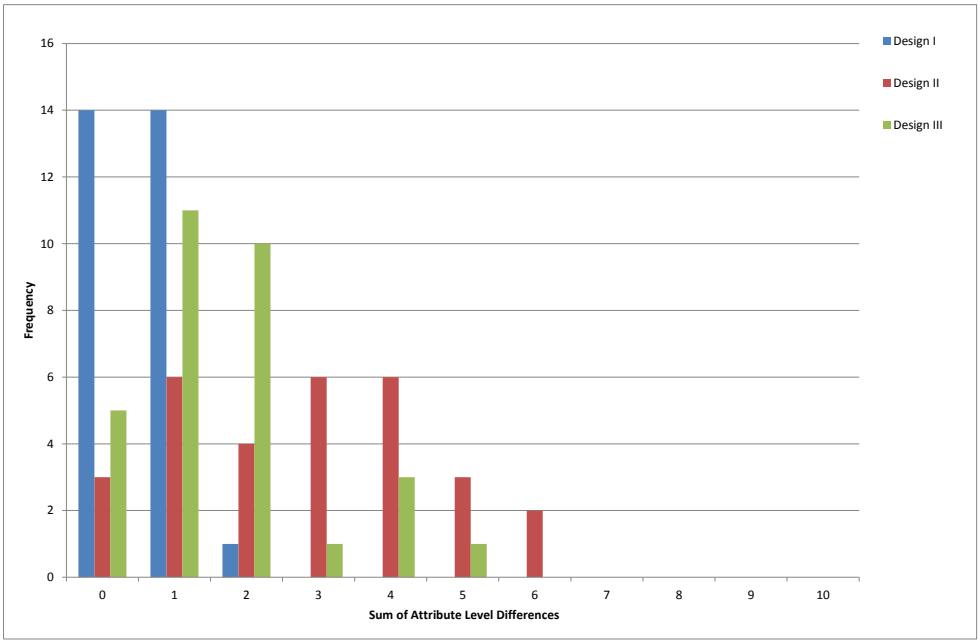

Figure 6.2: The absolute value of the sum of the attribute level differences between alternatives for Bayesian choice designs optimised using different prior distributions of the parameters: i.i.d U[0,1], non-identical U[0,1] and the TTO priors

To investigate the cause of this effect, we compare the choice tasks selected for the Bayesian design based on each prior distribution using the absolute utility sum differences that represents the absolute value of the sum of the attribute level differences of the alternatives within a choice task as shown in Figure 6.2. This is because this value reflects the distance between alternative within a choice task in terms of the mean utility, where larger utility sum difference indicates larger distance between the alternatives and their attribute levels in a choice task. For instance, a choice set of the AQL-5D health states 20033 and 14423 has an absolute utility sum difference of $|(2-1)+(0-4)+(0-4)+(3-2)+(3-3)|=6$, larger than health states that are more close to each other such as 20033 and 21123 which has a difference of one. In this case, therefore, health states with the same utility sum difference, such as choices in the AQL-5D choice sets $C_1 = \{01222, 01130\}$ and $C_2 = \{01402, 01013\}$, will have

similar distance between the alternatives in terms of utility.

Figure 6.2 indicates that designs II and III have more choices with large absolute differences, i.e. large distance between alternatives within choice tasks, than design I. Also, designs II and III give rise to choices with larger level differences between the less important attributes, while keeping the difference between the important attributes as small as possible. This is illustrated for design II in Figure 6.3, where the last two attributes, sleep and activities, typically have small level differences of one, and the other attributes such as the concern attribute typically have a larger difference of three and four.

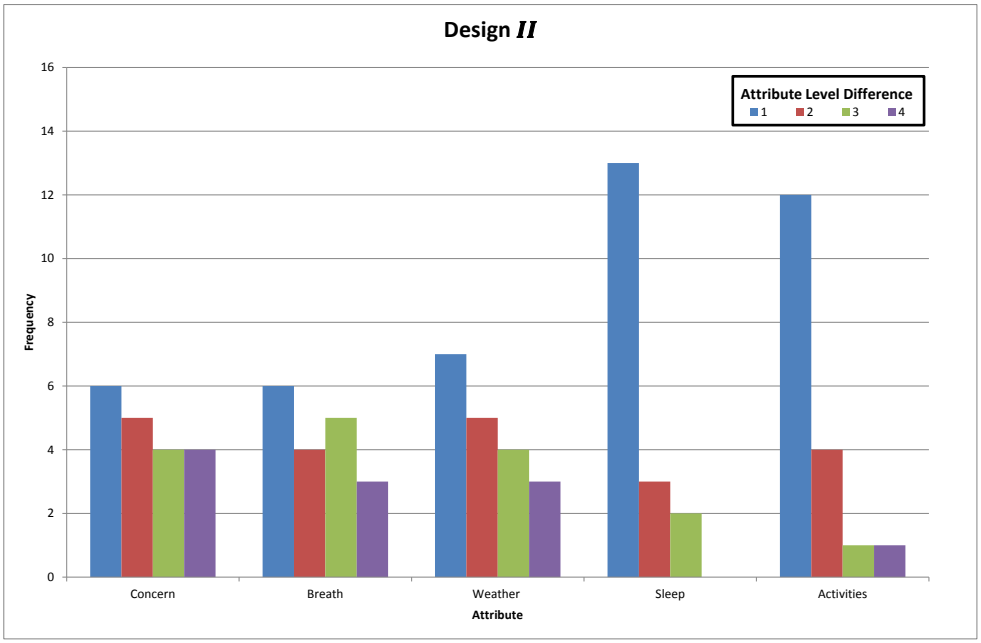

Figure 6.3: Attribute level differences between choices in the Bayesian design II which is optimised based on non-identical prior distributions of the parameters (level differences of zero that related to the constant attributes are not displayed here)

This is because, in prior 2 (the design prior), the sleep and activities attributes are more important than the remaining attributes. The design constructed based on identical prior distributions of the parameters typically has level differences of 1

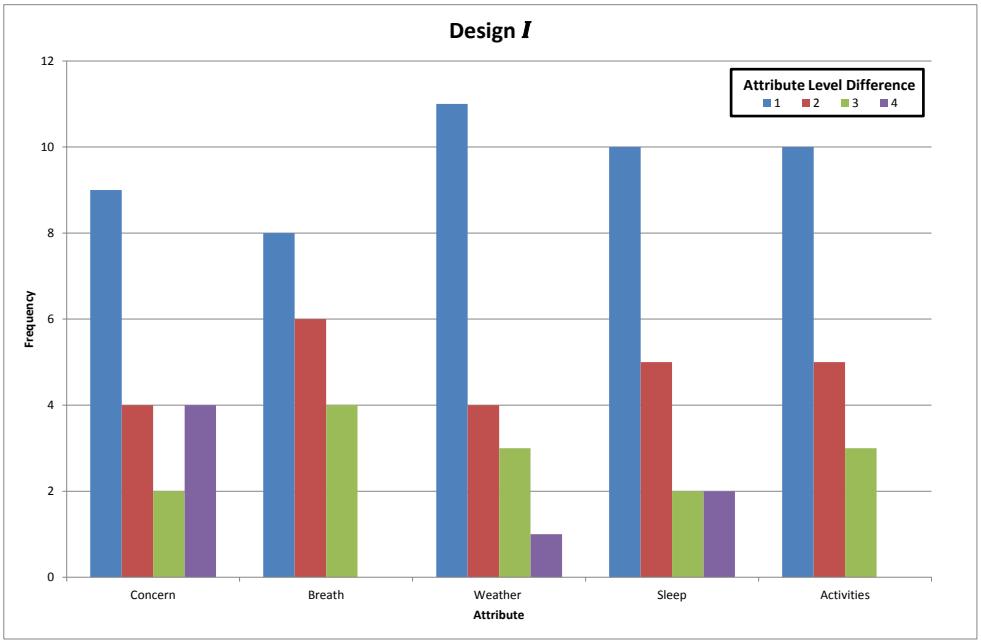

Figure 6.4: Attribute level differences between choices in the Bayesian design I that optimised based on non-identical prior distributions of the parameters (level differences of zero that related to the constant attributes are not displayed here)

between all the attributes, i.e. it reflects the small distances between the alternatives within a choice task, as shown in Figure 6.4 for design I optimised for the i.i.d U[0,1] prior of the parameters.

The investigation shows that poorly defined prior distributions might result in a large reduction of the design efficiency, particularly, when non-identically prior distribution for the parameters is used to generate the choice design, while the 'true' parameter values are identically distributed and vice versa.

6.4.2 Comparing Design Efficiency for i.i.d Priors of the Parameters

In this section, we conduct a similar investigation as in the previous section, but for different simple i.i.d priors for the logit model's parameters, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)$. We consider two sets of simple i.i.d priors. The first consists of distinct i.i.d prior of the parameters with a small and a large prior mean values of the parameters as follows:

- prior 1: $\beta_1, \ldots, \beta_{20} \overset{iid}{\sim} U[0, 0.1]$, and $\sigma \sim U[0, 0.1]$;
- prior 2: $\beta_1, \ldots, \beta_{20} \overset{iid}{\sim} U[0.9, 1]$, and $\sigma \sim U[0.9, 1]$.

The second set consists of prior distributions that provide values between 0 and 1 for the parameters but mainly differ in the shape of the distribution as follows:

- prior 3: $\beta_1, \ldots, \beta_{20} \overset{iid}{\sim} Gamma(1, 10)$, and $\sigma \sim Gamma(1, 10)$;
- prior 4: $\beta_1, \ldots, \beta_{20} \overset{iid}{\sim} Gamma(5, 15)$, and $\sigma \sim Gamma(5, 15)$.
- prior 4: $\beta_1, \ldots, \beta_{20} \overset{iid}{\sim} Beta(1, 10)$, and $\sigma \sim Beta(1, 10)$.

We generate Bayesian choice designs for the suggested priors in both sets, and compare the effect of the choice of the prior on the choice of Bayesian design based on the design criterion values, as described earlier in Section 6.4.1. The resulting designs from each set of priors are also compared to those generated based on the i.i.d U[0,1] prior as illustrated in Tables 6.2 and 6.3, respectively.

Table 6.2 illustrates that the choice of the Bayesian design is robust to particular choice of the uniform distributions, where misspecifying the prior distribution results in a small loss of the design efficiency unless the uniform distribution has large mean for the parameter values. For instance, if the 'true' prior distribution is U[0,0.1], using design II instead of design I produces an efficiency loss of 11% which is larger than the

Table 6.2: $\mathcal{D}^B_{\text{FIM}}$ design criterion values for different $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$-optimum choice designs evaluated at the design prior (shaded value) and the other suggested prior distributions together with the efficiency of the optimal design with respect to other designs

$\mathcal{D}^B_{\mathbf{s},FIM}$ ***optimal Designs***	***Design Prior***	***Evaluation Prior***		
		U[0,0.1]	U[0.9,1]	U[0,1]
I	U[0,0.1]	$2.35e^{-05}$*	0.0098	0.0047
II	U[0.9,1]	$2.64e^{-05}$	**0.0084***	0.0052
III	U[0,1]	$2.36e^{-05}$	0.0086	**0.0042***
$\mathcal{D}^B_{\mathbf{s},eff}$	-	89.01%	85.71%	89.36%
	-	99.57%	97.67%	80.77%

* The best criterion value in each column.

one produced when using design III (i.e. efficiency loss of 0.4%). This loss increases if the 'true' prior distribution is U[0,1], where using design II instead of design III results in an efficiency loss of approximately 19%.

Similar results are obtained when comparing the effect of the similar i.i.d priors of the parameters in the second set, i.e. G(1,10), G(5,15), Beat(1,10) and U[0,1] prior distributions as shown in Table 6.3.

Table 6.3: $\mathcal{D}^B_{\text{FIM}}$ design criterion values for different $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$-optimum choice designs evaluated at the design prior (shaded value) and the evaluation prior together with the efficiency of the optimal design with respect to other designs

$\mathcal{D}^B_{\mathbf{s},FIM}$ ***Optimal Designs***	***Design Prior***	***Evaluation Prior***			
		Gamma(1,10)	Gamma(5,15)	Beta(1,10)	Unif(0,1)
IV	Gamma(1,10)	**0.00030***	0.0016	0.00015	0.0047
V	Gamma(5,15)	0.00031	**0.0014***	0.00013	0.0043
VI	Beta(1,10)	0.00030	0.0016	**0.00014**	0.0047
VII	Uniform(0,1)	0.00030	0.0014	0.00013*	**0.0036***
$\mathcal{D}^B_{\mathbf{s},eff}$	-	96.77%	87.50%	93.33%	89.36%
	-	99.99%	87.50%	92.85%	97.67%
	-	99.99%	99.99%	92.85%	89.36%

* The best criterion value in each column.

Nevertheless, the impact of misspecifying the prior distribution for similar i.i.d priors is much more smaller than that observed earlier for the distinct i.i.d uniform priors of the parameters (Table 6.2). The maximum efficiency loss of 12% is produced with respect to Gamma(5,15) prior distribution when design IV or VI is used instead of design V.

To assess the robustness of these results to the choice of the starting designs, we replicate the analysis with fifteen different starting designs, as in Section 6.4.1. Figures 6.5 and 6.6 show the results of the robustness analysis.

The plots indicate a great consistency with the results shown in Tables 6.2 and 6.3, where misspecifying the prior distribution does not have a significant effect on the design criterion values, and the effect of the distinct prior distributions on the choice design is still larger than that observed for similar prior distributions. Also, the plots show some variation on the performance of particular prior distribution with respect to other Bayesian designs optimised based on different prior using different starting designs, where some of these designs perform as well as or better than design optimised with respect to the correct prior. Thus, in this case, the small effect of the prior distribution on the choice of the Bayesian designs might be related to the choice of the random starting design used in the optimisation procedure rather than the actual choice of the i.i.d prior for the parameters.

6.4.3 The Sensitivity of Bayesian Choice Design to the Prior Distribution of the Scale Parameter

In the previous investigations, the same prior distributions are assumed for both the preference parameters $\boldsymbol{\beta}$ and the scale parameter of the random error variance, σ. In this section, we examine the impact of the choice of the prior distribution of σ only on the choice of the Bayesian optimal design. This is done by generating Bayesian choice designs for different choices of the prior distribution of the scale parameter σ, while

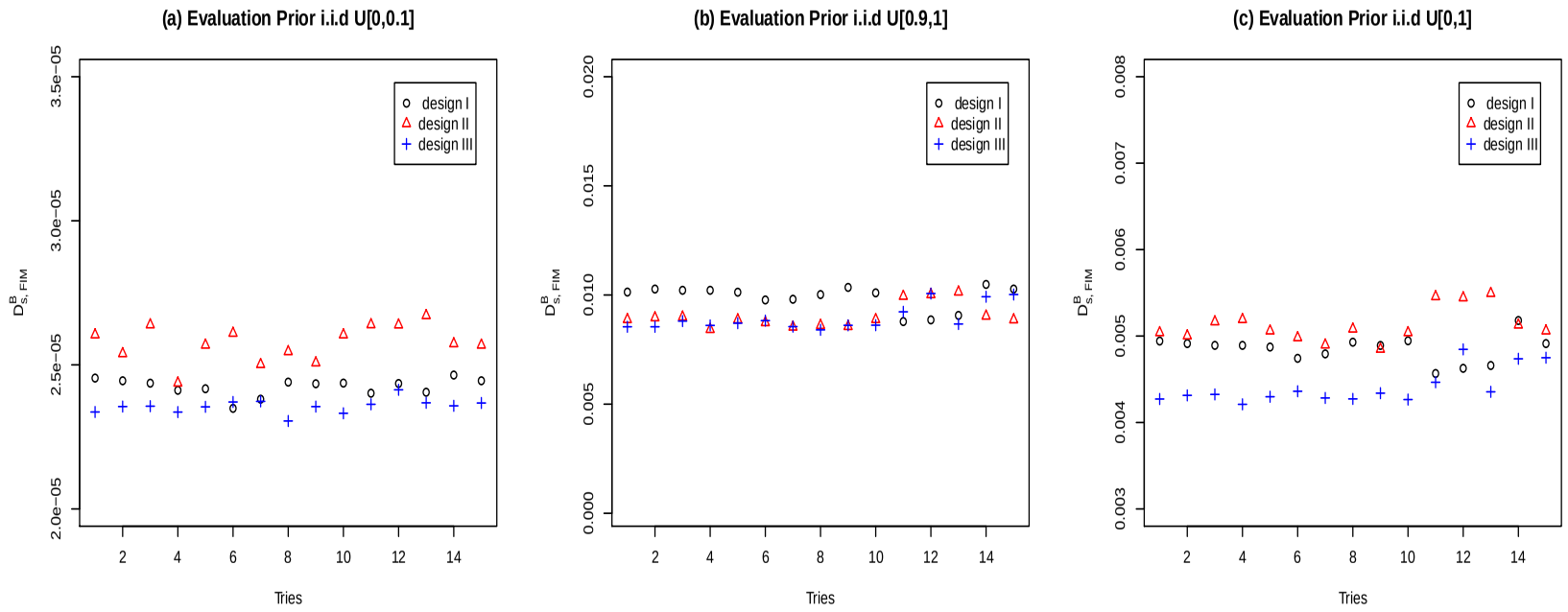

Figure 6.5: Panel plots of the $\mathcal{D}^B_{\mathbf{s},\text{FIM}}$ design criterion values computed for different prior distributions with respect to different Bayesian designs, where designs I, II and III are the corresponding Bayesian choice designs obtained based on the U[0,0.1], U[0.9,1] and U[0,1] priors, respectively

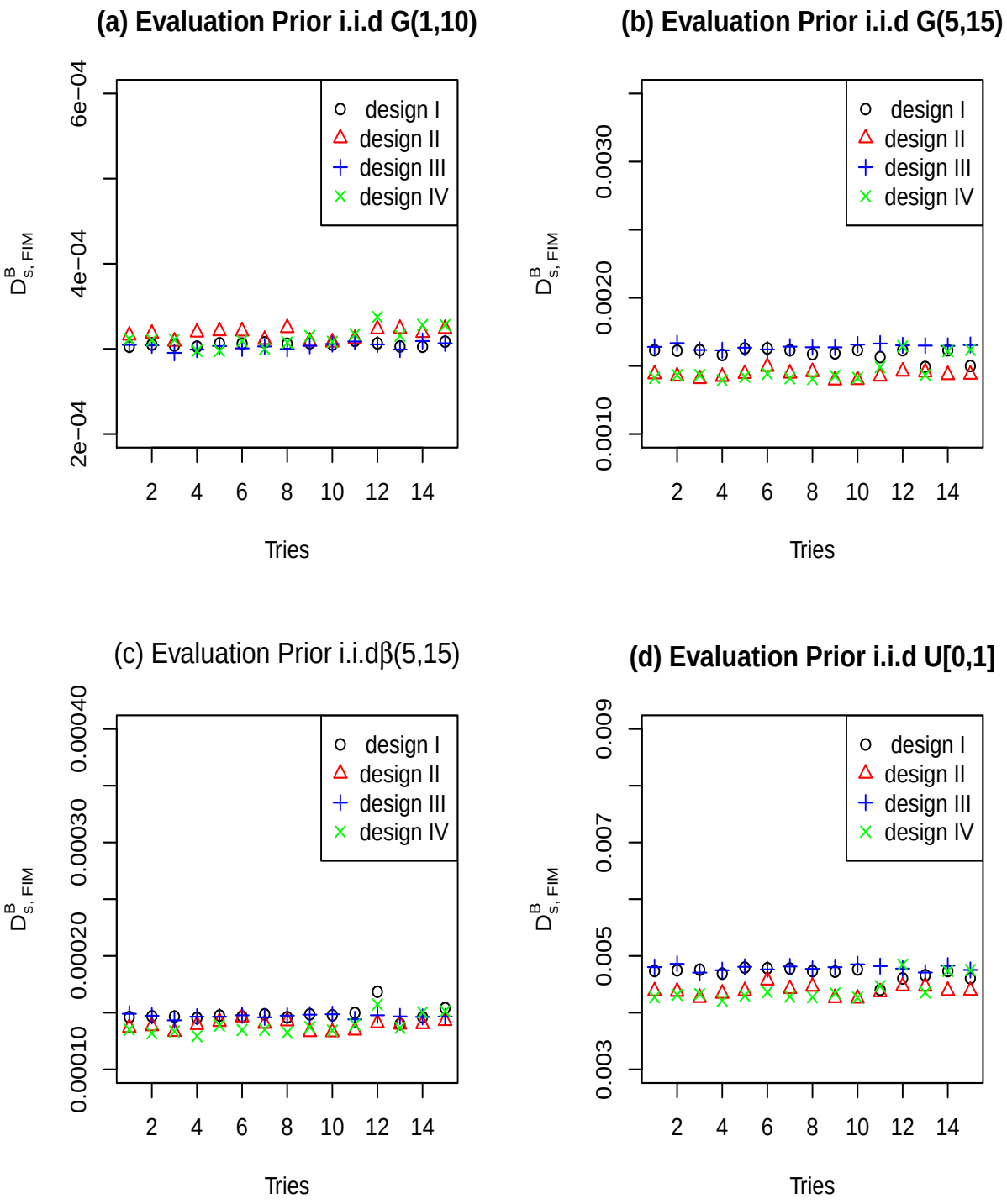

Figure 6.6: Panel plots of the $\mathcal{D}^B_{\text{s,FIM}}$ design criterion values computed for different prior distributions with respect to different Bayesian designs, where designs I, II, III and IV are the corresponding Bayesian choice designs obtained based on the Gamma(1,10), Gamma(5,15), Beta (1,10) and the U[0,1] priors, respectively

fixing the prior distributions of the preference parameters, $\boldsymbol{\beta}$. We then compare the efficiency of the resulting designs using the design criterion values as described in the earlier sections.

For illustration, we assume that $\beta_1, \ldots, \beta_{20} \overset{iid}{\sim} U[0, 0.1]$, and allow the prior distribution of the scale parameter σ to take different uniform distributions with relatively a larger value compared to the value of the preference parameters, as well as the TTO prior of the scale parameter as follows:

- prior 1: $\sigma \sim U[0, 0.1]$
- prior 2: $\sigma \sim U[0, 1]$;
- prior 3: $\sigma \sim U[0.5, 1]$;
- prior 4: $\sigma \sim U[0.9, 1]$;
- prior 5: $\sigma \sim U[2, 3]$;
- prior 6: $\sigma \sim \text{TTO}_\sigma$ prior;

where TTO_σ represents the posterior distribution of the scale parameter obtained form Bayesian analysis of the TTO data. We then evaluate the resulting Bayesian choice designs for these priors with respect to each prior distribution for which the design was not optimised, and compare these values to those obtained using a design generated based on the same prior for both the preference and the scale parameters, i.e. the U[0,0.1] prior, as shown in Table 6.4.

To study the effect of misidentifying the prior distribution of the scale parameter, σ, on the choice of Bayesian design itself, we compare the criterion values for particular prior distributions with respect to different choice designs. Considering the design criterion values to the U[0,0.1] evaluation prior, it can be seen that Bayesian designs II to V, i.e. designs optimised with respect to prior distributions with a large expected

Table 6.4: $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$ design criterion values for different $\mathcal{D}^B_{\mathbf{S},\text{FIM}}$-optimum choice designs evaluated at the design prior distribution of the variance parameter (shaded value) and the other proposed prior distributions

$\mathcal{D}^B_{\mathbf{S},\textbf{FIM}}$ ***Optimal Designs***	***Variance Design Prior***	***Variance Evaluation Prior***					
		U[0,0.1]	U[0,1]	U[0.5,1]	U[0.9,1]	U[2,3]	TTO_σ
I	U[0,0.1]	$2.35e^{-05*}$	0.0013	0.0023	0.0034	0.0240	$2.65e^{-05}$
II	U[0,1]	$2.75e^{-05}$	**0.0012**	0.0023	0.0034	0.0236*	$3.03e^{-05}$
III	U[0.5,1]	$4.20e^{-05}$	0.0006	**0.0026**	0.0035	0.0244	$4.39e^{-05}$
IV	U[0.9,1]	$4.03e^{-05}$	0.0004*	0.0026	**0.0032**	0.0249	$4.27e^{-05}$
V	U[2,3]	$3.45e^{-05}$	0.0006	0.0020*	0.0031*	**0.0246**	$3.66e^{-05}$
VI	TTO_σ	$2.44e^{-05}$	0.0014	0.0023	0.0037	0.0247	$\mathbf{2.60}e^{-05*}$

* The best criterion value in each column.

value for σ, reduce the design efficiency by approximately 15% to 44% compared to design I (design optimised for the correct prior, U[0,0.1]).

This reduction is observed because prior distributions with large expected values for σ generate more choices with a large absolute differences compared to designs optimised for prior distributions with smaller expected values of σ, such as the U[0,0.1] and the TTO prior distributions, as illustrated in Figure 6.7. These choices with a large attribute level digfferences would maximise the vector of the attribute level difference between alternatives, $(\boldsymbol{x}_{1s} - \boldsymbol{x}_{2s})$, but produce less balanced choice probabilities for these alternatives and hence reduce the efficiency of the choice design under the U[0,0.1] and the TTO prior distribution.

Also, a larger expected value for the scale parameter might result in a dominant choice task if no constraints on the dominant alternatives are specified in the exchange algorithms. Looking more closely at designs III to V, we observe that the number of dominant choice sets in these designs increases as the expected value of the scale parameter increases. This follows since the logit choice probability is a function of the mean utility difference and the scale parameter, as in

$$P_{1s} = \frac{1}{1 + \exp\left(\frac{\boldsymbol{\beta}(\boldsymbol{x}_{1s} - \boldsymbol{x}_{2s})^T}{\sigma}\right)},$$

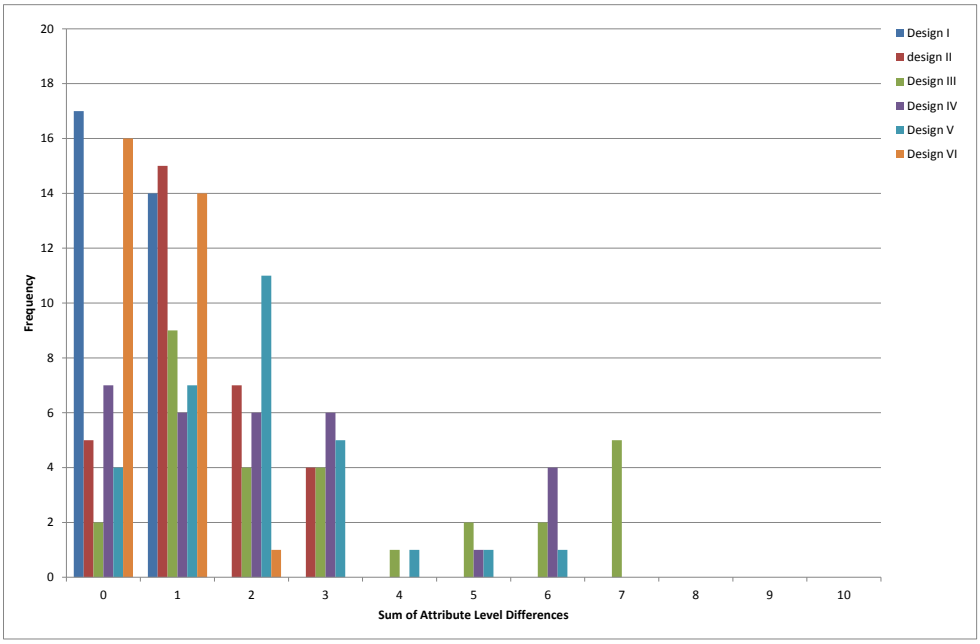

Figure 6.7: The absolute value of the sum of the attribute level differences between alternatives for Bayesian choice designs optimised using different prior distributions of the scale parameter presented in Table 6.4

and increasing the value of the scale parameter relative to the preference parameter values improves the balance of the choice probabilities of the alternatives irrespective of the actual difference between their attribute levels. Therefore, optimising the design criterion using the exchange algorithm cannot distinguishes between dominant and non-dominant alternatives as $\sigma \to \infty$. In this case, dominant choices are more likely to be selected as they improve the value of the design criterion. For instance, the U[0.9,1] prior distribution of σ provides similar mean utility values and, hence, balanced choice probabilities for the dominant alternatives in the choice task $C_s = \{00000, 44444\}$, which is unrealistic.

Therefore, a poorly defined prior distribution of the scale parameter, e.g. a prior distribution with a large expected value for the scale parameter relative to the preference parameters, produces choices with large gaps between attribute level (i.e. one-sided

choices), which impact badly on design efficiency. Therefore, it is worthwhile to obtain a more informative prior of the scale parameter, σ, in order to avoid one sided options and generate a more efficient Bayesian choice design.

6.5 Summary and Discussion

In this chapter, we used two different methods to assign prior distributions for the unknown logit model parameter vector $\boldsymbol{\theta}$: simple prior distributions constructed based on the available information about the unknown parameters, and prior distributions derived from Bayesian analysis of a previous study presented in Chapter 4. We then carried out an analysis to study the effect of the choice of the prior distribution of the parameters on the choice of Bayesian designs.

The analysis showed that overall the choice of the prior distribution influences the choice of Bayesian design, and consequently the design efficiency. Nevertheless, the amount of efficiency loss with respect to particular prior distribution depends on the type of prior used to optimise the choice design. Thus, if in the 'true' prior the parameters are not identically distributed, i.e. some attributes are more important than the others, then designs optimised for i.i.d priors for the parameters may reduce the design efficiency by more than a half relative to the choice design optimised based on the 'true' prior distribution and vice versa. This reduction is due to the fact that non-identical prior distributions, as opposed to identical prior distributions, prefer more choices with larger attribute level differences while keeping the level of the most important attributes similar and increasing the differences between the less important attributes.

The analysis also suggest that if i.i.d priors are appropriate for the parameter values (i.e. when the attributes are equally important to respondents), the precise choice of the prior is less important. We also investigated the effect of the prior distribution of the scale parameter, σ, on the choice of the Bayesian designs. The analysis showed that a poorly defined prior distribution for this parameter reduces the efficiency of the choice design. In particular, increasing the values of this parameter relative to the values of the preference parameters will increase the number of dominant choices in the design if no constraint is specified in the exchange algorithm. This is because the logit choice probability depends on the value of the scale parameter, and larger values for this parameter relative to the preference parameters balance the choice probabilities of the alternatives within a choice task irrespective of their actual attribute level differences.

Overall, our illustration study shows that the choice of Bayesian designs is robust to particular choices of identical prior distributions, whereas switching between these prior and non-identical prior distributions of the parameters results in significant loss in the design efficiency. In our analysis, the TTO prior distribution was the most appropriate prior of the parameters, as it was based on previous data. This prior favours some attributes as more important, and hence prefers choices with larger attribute level difference between the less important attributes while keeping the levels of the most important attribute similar. The analysis showed that this prior performs badly in terms of Bayesian choice designs constructed based on the i.i.d priors, and more worse with respect to Bayesian designs optimised for the non-identical prior that mimics the TTO prior in terms of presenting the most important attributes but with larger expected values for the parameter associated with these attributes. Therefore, poorly specified priors for the parameters really do matter, they might results in a large loss of design efficiency.

This efficiency loss requires more respondents to perform the choice experiment-possibly twice the number of respondents used in the choice experiment - to return the same level of precision in estimating the preference parameter values as in Bayesian

design optimised based on the correct prior distribution. The number of respondents is usually a significant matter in many health evaluation studies, where the number of participants is limited. This is because many health researches have research-sources constraints or study rare health conditions that limit the number of sample size between 100-300 respondents (Marshall et al., 2010). Therefore, we suggest that it is worthwhile for experimenter to work to derive more appropriate prior distributions, particularly in terms of identifying the most and least important attributes under study, as well as the relative importance of these attributes, to generate a more efficient choice design instead of just increasing the number of participants. The prior of the parameters can be obtained from previous study as in our case study. Nevertheless, in the case where there is no data available for the underlying classification system, then one can derive an appropriate prior by:

1. eliciting experts' prior beliefs about the relative importance of the attributes under study;

2. mapping the data from other classification systems with pre-existing health state values (e.g. EQ-5D system) to the underlying case study (e.g. the AQL-5D system). This method allows us to predict the utility scores for the AQL-5D health states, and hence estimate the preference parameters associated with attribute levels of this classification system using different regression techniques. These techniques require specifying the classification system to map from and map to (see Franks et al., 2004; Gray et al.,2006; and Ara and Brazier, 2008, for more details about these techniques)

Chapter 7

Conclusion

In this chapter, we discuss the main findings of the thesis. This thesis has aimed at developing a efficient methodology to construct efficient choice designs for valuing health state utilities using the latest advanced work in the optimal design theory. To achieve that, the main experimental design issues and considerations to generate choice experiments for health evaluation studies were firstly identified. Then, related advanced work in optimal design literature, particularly Bayesian optimal design, was reviewed to investigate its ability to improve the choice design for valuing health.

In this chapter, we begin by summarising the contributions previous chapters make to our objective, and then in Section 7.2 discuss the main findings of each chapter. Sections 7.3 and 7.4 provide the limitations of the thesis and direction for further work, along with recommendations based on the findings of the thesis, respectively.

7.1 Summary of the Thesis

The main concepts for health economics evaluation are reviewed in Chapter 2 to identify relevant background knowledge. In particular, the chapter presented the use of QALYs in making a decision for allocating limited financial health resources, and different classification systems to describe health outcomes/health states (e.g. EQ-5D and AQL-5D systems); as well as different techniques to measure the utility values of these states, i.e. the 'Q' part of the QALY. The main issues with each technique were discussed, with a view to identifying new evaluation methods that give more reliable health state utility values, with more emphasis on the discrete choice experiments (DCEs) technique and their modelling.

The DCE method seems to be a promising alternative for the direct valuation methods, which would reduce survey administration times and efforts for collecting data. This is because DCEs are relatively easy to comprehend and administrate, and, as opposed to the TTO method, respondents usually do not require face to face interview to perform such choice task. However, the literature review in Chapter 3 illustrated that more work is still required to improve the choice data collected from this technique and consequently the estimated utility values, particularly in terms of the experimental design used to select choices presented to respondents and group them efficiently into choice tasks, as well as anchoring the utility values within the required QALY scale (0–1 scale). In Chapter 3, therefore, advanced methodologies for deriving efficient DCEs in various areas outside health economics were reviewed, particularly for Bayesian optimal choice designs, in order to construct better experimental designs for valuing health state utilities. Constructing Bayesian optimal design for discrete choice models, unlike linear models, depends on the unknown model parameters in the information

matrix, and hence so does the optimal design criterion. Therefore, a prior distribution of the unknown model parameters is required to construct an efficient choice design.

In Chapter 4, two datasets for the AQL-5D health states using time trade-off (TTO) and pairwise choice experiment methods were analysed using classical and Bayesian approaches to obtain prior information about our model parameters, and demonstrate the effect of the type of data on estimating the health states utilities. In particular, a Bayesian analysis was performed on the TTO data to provide a prior distribution of the unknown model parameters for generating Bayesian optimal designs for the same classification system. Also, a Bayesian analysis was used to compare the uncertainty in the mean utilities produced by the TTO and discrete choice models. The result of the analysis illustrated that DCE data produces slightly higher uncertainty in the estimated mean utility values compared to TTO data. However, this might be related to the number of respondents and health states used in each method as well as the way the choice design is constructed, where more observations is obtained under the TTO exercise than for the DCE task.

Therefore, we suggested using more sophisticated methods to improve the quality of the collected choice data and hence the final results – in particular, using Bayesian optimal designs that account for uncertainty in the model's parameters and the utility values by incorporating the prior information in the phase of constructing the choice design. Chapter 5, therefore, provided the optimal design criterion for our model, here a logit model, and then investigated the possibility of using available design software such as SAS, JMP or Ngene to construct Bayesian pairwise choice design for valuing AQL-5D health states within the QALY scale. We have not managed to generate Bayesian optimal choice designs using these programs, as they are limited in handling our design problem, particularly in terms of including the death state in the choice design to anchor the utility values and optimising the correct design criterion.

Thus, we proposed our design algorithm that is based upon combining and modifying different advanced search algorithms available in the Bayesian design literature

in such a way that it handles our design considerations. The algorithm is particularly developed to construct Bayesian pairwise choice designs for the logit model based on the Bayesian $\mathcal{D}_{\mathbf{S}}$-optimality criterion that accounts for the inclusion of death. An application for our algorithm using the AQL-5D case study and logit model is provided in Chapter 5.

Chapter 5 also provided a method to simplify the choice designs by holding some attributes constant in the choice tasks, to reduce the error in the respondents' choices (i.e. increase the response efficiency). This simplification results in a reduction of the statistical efficiency of the choice design, while on the other hand it increases the response efficiency and consequently the reliability of the choice data.

The performance of our Bayesian choice designs might depend on the prior distribution chosen to construct the designs. Therefore, the effect of the prior distribution on the actual choice of the Bayesian choice designs was investigated in Chapter 6. The chapter compared different Bayesian designs optimised based on different type of prior distributions based on their optimality design criterion values, i.e. the efficiency of particular design compare to the other. The analysis illustrated that misspecifying the prior distribution may badly affect the efficiency of the choice design.

7.2 Discussion of the Main Findings

In health economics, health states utilities are evaluated using different techniques. Chapter 2 illustrated that contamination of the utility values produced by the direct valuation methods, such as time trade-off (TTO) and standard gamble (SG) techniques, by non-health factors, together with their complexity, increase the interest in using the

discrete choice experiments (DCEs) method. This method is more straightforward than, for example, the SG method, and reflects more accurately the type of decision that individuals make every day.

Nevertheless, there are still some issues related to using the DCE technique within a health evaluation context, particularly in terms of the experimental design for constructing the DCE, which can be summarised as follows.

- Health state utility values inferred using discrete choice data are not directly anchored on the death and perfect health scale (0–1 scale) required for the QALY calculation, and hence cannot be used directly in cost-utility analysis (CUA). In this thesis we considered including the death state comparison in the choice design to anchor the health state utilities, as in Brazier et al. (2009). Since other anchoring methods either depend on other valuation techniques rather than the DCE, such as the TTO method, or complicate the design choice problems by including the survival attribute in the design. This is conflict with our aim of establishing an efficient DCE as an alternative for the direct valuation methods and simplifying the choice design problem. Thus, experimental design should account for the inclusion of the death state in the choice designs.

- The construction of efficient choice design requires many constraints to improve the collected choice data, and consequently the reliability of the preference parameter estimates and the utility values. In particular, it is necessary to impose constraints on attribute level combinations to avoid dominant and implausible health states defined by a classification system, since they reduce the efficiency of the choice design.

- The non-linear nature of the discrete choice models complicates the design problem, and usually standard designs such as orthogonal designs are not suitable.

- The construction of the choice design depends on the values of the unknown model's parameters. Therefore, usually prior information about the parameters

is required, and a Bayesian approach is sought to generate the design.

Generating an efficient choice design for valuing health states therefore requires an algorithmic experimental design to account for the design considerations. But the results of the literature review in Chapter 3 showed that most of the choice designs used for valuing health state utilities are based on either orthogonal array designs or the optimal design principles – i.e. level balance, orthogonality, minimal overlap and utility balance – developed in Huber and Zwerina (1996). Also, most designs ignore the dependency of the choice designs on the model's parameters by assuming zero prior point estimates for the preference parameters instead of using a Bayesian approach. This might result in dominant and implausible choices, as in the level balanced design (LBD) generated for the AQL-5D system, which reduces the efficiency of the choice design. Since respondents might have difficulty in evaluating implausible health states, that complicates the choice task and increases the error variance, and in dominant choice tasks all respondents will choose the dominant state and hence such tasks would provide no valuable information about the preference parameters.

Having identified the main design considerations and issues for valuing health states utilities, we considered improving the methodology for constructing the choice design based on the latest optimal Bayesian experimental design method in the design literature. The main key features of our design algorithm are as follows:

- it uses the correct design criterion; i.e. it consider optimising the design with respect to the parameters of interested and accounts for the including of the death state in the choice model, to optimise the choice design;
- it is flexible to any choice of prior distributions and any design constraints; and
- it considers simplifying the choice task to reduce the error in the respondents' choices.

Applying our design algorithm to generate a Bayesian pairwise choice design for the

AQL-5D case study illustrates a substantial improvement over the LBD, particularly when the true parameter values are not zero (which is almost certainly the case in most choice studies). This is because our design algorithm accounts for the uncertainty in the model parameters by incorporating a priori information when constructing the choice design. Also, the use of the coordinate-exchange algorithm together with a suitable prior distribution for the preferences parameters accounts for the separation between alternatives in a choice task, and this results in more efficient choice questions and eliminates dominant choices from the choice designs. This improves the efficiency and the information collected from the choice design, and consequently the estimated preference parameter and utility values.

Simplifying the Bayesian choice design introduces some reduction in the design efficiency. But this reduction is not remarkable compared with the detrimental effect that non-simplified choice design might have on the respondents' choices and consequent reliability of the utility values. This is because respondents usually violate the compensatory assumption when varying many attributes in alternatives within a choice task, and hence make their choices based on trading off between a subset of the attributes instead of considering all the attributes under study, which might affect the final results. Therefore, to obtain a reliable assessment of health state utilities, we recommend using the simplified Bayesian choice design, particularly for a design with a large number of attributes and attribute levels.

The choice of the prior distribution might have an impact on the choice of the efficient Bayesian design. Our illustration study, which compared different Bayesian pairwise choice designs optimised with respect to different prior distributions, indicated that misspecifying the prior distribution might have a substantial effect on the Bayesian designs, particularly when switching between identical and non-identical prior distribution of the preference parameters. In case all attributes are equally attractive to respondents, the choice of the identical prior distribution of the parameters is less important as it has a smaller effect on the choice of the efficient Bayesian design. Also,

poorly defined prior distributions for the scale parameter of the random error, e.g. a prior with a large value for the scale parameter relative to the preference parameter values, produces a less efficient choice design. This is because such a choice of prior distribution prefers a large distance between alternatives within a choice task irrespective of the actual difference between their attribute level, and that might result in dominant choices if no constraint is specified in the exchange algorithm.

A large design efficiency loss requires a large number of respondents to perform the choice experiment in order to return the same level of precision in estimating the preference parameter values as in Bayesian design optimised with respect to the correct prior distribution. This might mean more than twice the number of respondents being used in the choice experiment, particularly when switching between identical and non-identical prior distributions. Therefore, it would be worthwhile for an experimenter to derive appropriate prior distributions that reflect the preference of the attributes under study from the point of view of the respondents (e.g., most and least important attributes), and use this prior information to generate more efficient choice designs instead of increasing the number of participants.

7.3 Limitation and Further Work

In this thesis, the Bayesian optimal choice designs were based on the Bayesian $\mathcal{D}_{\mathbf{S}}$-optimality criterion assuming a large sample size (i.e. respondents to value the choice design). The calculation of what sample size is required for a particular study is one problem worth studying in the Bayesian optimal designs, particularly for health evaluation studies where usually the number of participants is limited.

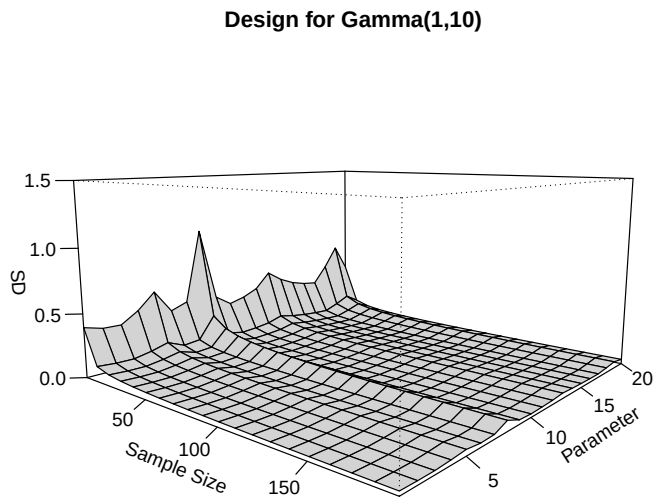

Figure 7.1: Standard error for the preference parameter associated with the attribute level difference in the AQL-5D system estimated using the FIM of the Bayesian design optimised for Gamma (1,10) prior distribution

The specification of the required sample size depends on the aim of the experiment, which is usually to estimate the parameters with high reliability. Our Bayesian designs minimise the $\mathcal{D}_{\mathbf{S}}$-optimality design criterion, which translates into minimising the volume of the posterior credible ellipsoid of the unknown preference parameters associated with the attribute level differences. However, this might result in less reliable estimates for some parameters compared to the other (i.e. larger standard error for some parameters). For instance, Bayesian $\mathcal{D}_{\mathbf{S}}$-optimum choice design constructed based on the Gamma (1,10) prior distribution provides larger standard error for the parameter associated with level 4 of the short of breath, β_8, as shown in Figure 7.1.

Therefore if experimenters are interested in estimating all the parameters with a high precision of, say, 0.05, then a preliminary solution to specify an appropriate number of sample size would be to set the maximum sample size required to increase the level of the precision for the less reliable parameters (i.e. β_8) as the minimum

boundary for the total number of sample size required (e.g. 150 respondents or more in this case).

Another suggestion would be to modify the choice design in such a way that it gains more information about the less reliable parameters and, hence, minimises the sample size required, as suggested in Rose and Bliemer (2013). However, this may require sacrificing some information about the other parameters. Thus, we encourage further research to develop more sophisticated methods to specify and minimise the sample size required to estimate the parameters with high reliability using Bayesian optimal choice designs.

In health economics, decision makers are more interested in increasing the reliability of decisions made between treatments, which depends on the reliability of the differences between the health state utilities. Thus it would be worth investigating and specifying the sufficient sample size required to detect particular differences between utilities, which represent treatment effect, as cost-effective within a certain level of precision, using our Bayesian choice designs and with respect to different prior distributions. Indeed, a further research of interest would be deriving a new optimal design criterion that aims to optimise the choice design so that the variance of a particular health difference is minimised, and then compare how the design efficiency and sample size required might be affected, compared to the Bayesian $\mathcal{D}_{\mathbf{S}}$-optimum design. In addition, it would be of interest to develop a prediction design criterion, such as the $\mathcal{V}$-optimality criterion, for predicting the mean utility value within the required QALY scale, since it is a key of activity in health economics.

As in our study, Bayesian choice designs involve many choice questions than can be performed by a single respondent (here 32 choice tasks); respondents typically cannot evaluate more than 12–20 choice questions before they become exhausted and start to provide less reliable choices (Johnson and Orme, 1996). Therefore, we suggest that a number of choice questions from the entire choice designs are randomly assigned to each respondent (e.g., 8 choice questions including the death comparison). Another

method to reduce the number of choices evaluated by each respondent is to allocate the choice questions into different blocks of equal size and then randomly assign each respondent to these blocks, or use the balanced incomplete block design approach that has been shown to yield suboptimal designs. However, an efficient method is needed to block the choice design such that it does not reduce the design efficiency.

7.4 Main Recommendations for Practical Applications

To increase the statistical efficiency of the choice design given a particular sample size, our results suggest using more appropriate prior distribution for the unknown model parameters, particularly when the attributes under study are not equally important. In practice, of course, the choice designs are constructed before observing the data in the field, where eliciting expert's prior beliefs about the relative importance of the attributes under study is important in this case. Nevertheless, obtaining experts judgements about these quantities might be an expensive and difficult task in practice. Therefore, one could use a prior that, at least, reflects the relative importance of each attribute from respondents' perspectives, i.e. the most and least important attributes under study. Also, using such a prior distribution together with an efficient search algorithm, such as the coordinate-exchange algorithm used in our approach, dominant choices can be eliminated from the final design without the need to impose many constraints in the design algorithm.

Additionally, though simplifying the choice design will result in slightly less sta-

tistical efficiency for a given simple size, we suggest using the simplified design in practice, particularly for a large design with a large number of attributes and attribute levels. This is because simplifying the choice tasks might reduce the error in respondents' choices, and force respondents to trade-off between attributes even if a dominant attribute exists. This then could increase the reliability of the choice data and consequently the estimated values for the preference parameters and the utilities, as the overall precision of these quantities depends on balancing both the statistical and response efficiencies (Johnson et al., 2013).

Appendix

A.1 Posterior Distributions

The posterior distribution of the preference parameters, $\boldsymbol{\beta}$, and the scale parameter σ obtained from the TTO data and Gamma(1,10) prior distribution are presented in this section.

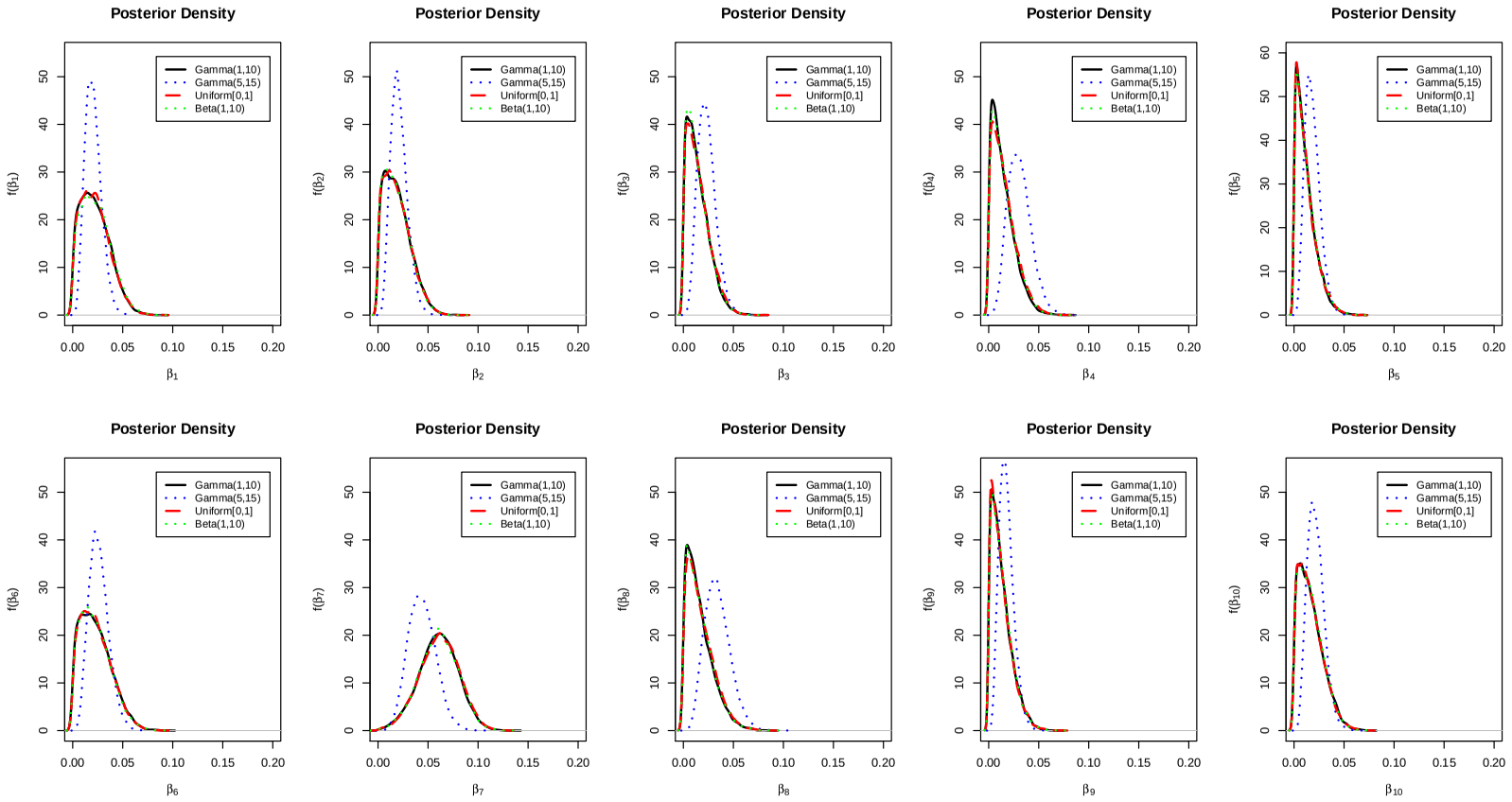

Figure 2: The posterior distributions of the preference parameters associated with the attribute level differences in the AQL-5D classification system

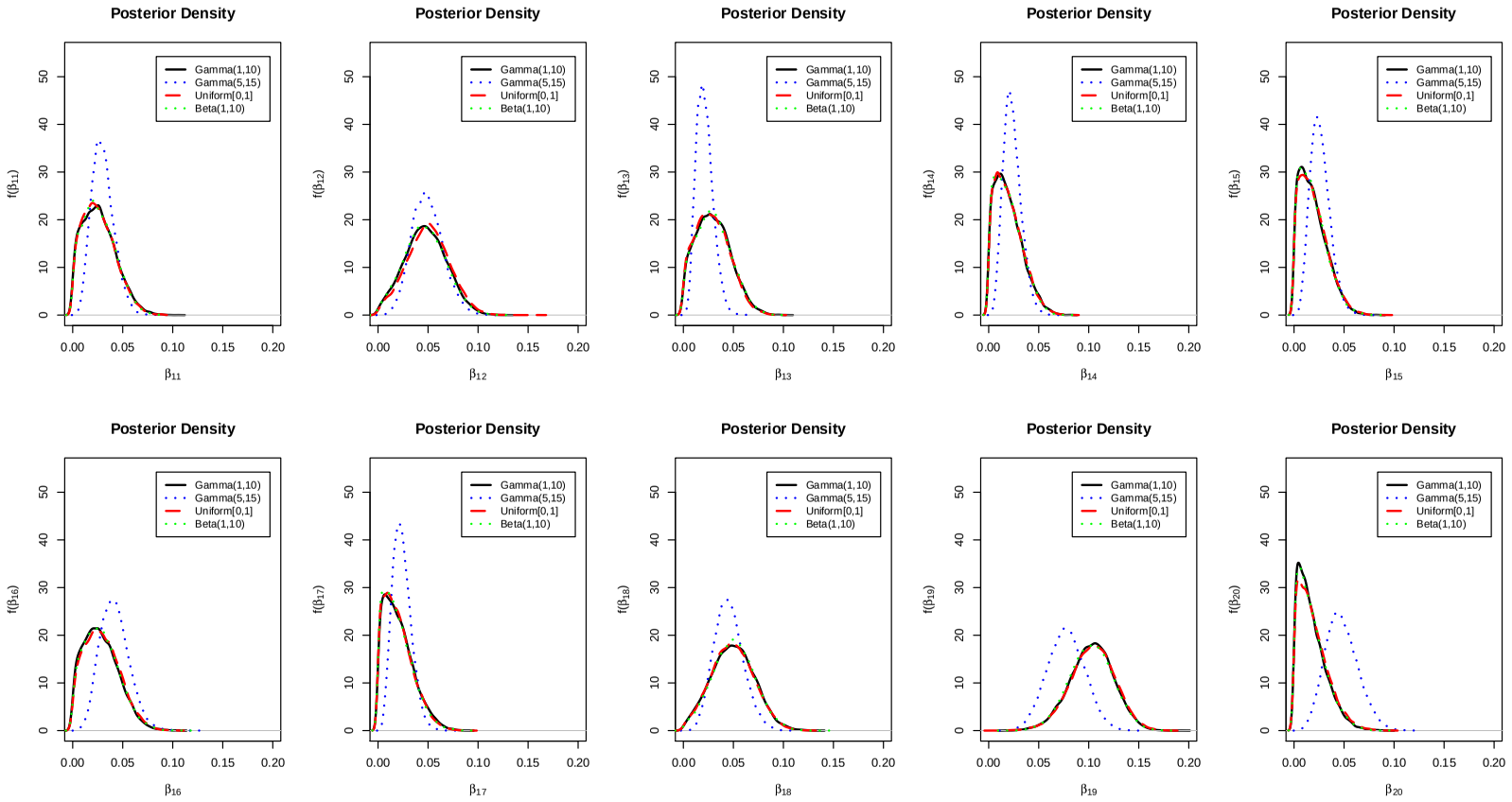

Figure 3: The posterior distributions of the preference parameters associated with the attribute level differences in the AQL-5D classification system

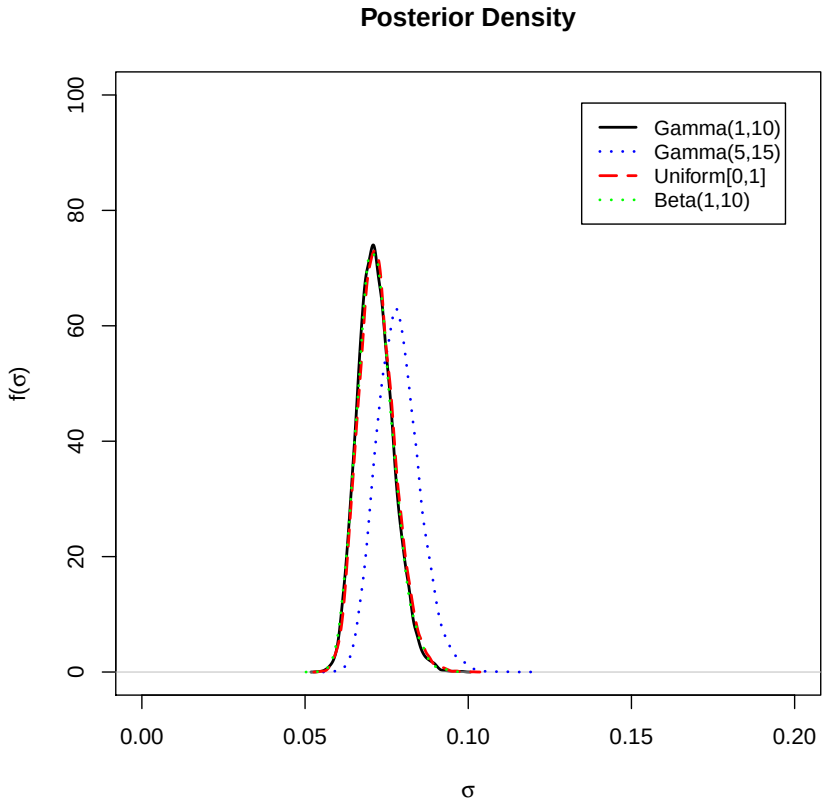

Figure 4: The posterior distribution of the scale parameter, σ, obtained from the TTO data and Gamma(1,10) prior distribution

A.2

Discrete Choice Designs

A.2.1 DCE for Asthma Health States Using Huber and Zwerina (1996) Approach

Table 1: DCE for asthma health states together with the total number of respondents evaluate each choice set, N, and the number of respondent select the first state over the second state in each choice set,n

Choice Set	Alternative I					Alternative II					N	n
1	3	4	2	0	3	2	1	3	3	1	39	7
2	2	0	3	0	4	1	2	1	2	0	39	4
3	2	4	4	0	3	4	0	1	4	4	38	17
4	1	2	1	0	1	4	3	2	2	0	39	33
5	2	1	0	2	1	4	0	4	3	3	38	35
6	1	3	0	3	0	4	4	1	2	1	38	29
7	4	3	0	1	1	1	0	3	2	2	35	15
8	2	1	4	4	3	4	3	3	0	0	35	11
9	3	1	4	2	4	2	2	2	4	2	35	4
10	0	4	4	3	0	1	0	0	1	3	35	16
11	1	4	2	3	4	3	3	4	4	2	34	13
12	2	4	2	2	3	4	1	3	0	4	34	22
13	3	1	3	3	4	2	0	4	1	0	51	10
14	1	4	4	4	1	2	2	1	3	3	52	19
15	1	3	2	4	4	0	0	0	2	0	52	4
16	3	4	0	3	2	0	1	2	0	0	52	3
17	0	2	0	4	4	3	0	2	3	1	52	11
18	3	2	3	4	0	0	3	1	0	2	52	12
19	0	1	2	1	2	2	3	1	2	4	42	38
20	0	0	0	4	1	4	1	3	1	2	42	26
21	4	1	0	0	0	0	3	3	1	3	42	39
22	3	2	2	1	4	4	4	3	4	3	42	34
23	3	4	1	1	0	4	2	4	2	1	42	34
24	2	2	3	2	1	3	3	2	3	2	42	40
25	4	4	4	4	4	-1	-1	-1	-1	-1	37	29
26	4	4	4	4	4	-1	-1	-1	-1	-1	35	25
27	4	4	4	4	4	-1	-1	-1	-1	-1	52	41
28	4	1	0	0	0	-1	-1	-1	-1	-1	42	41
29	-1	-1	-1	-1	-1	3	3	2	4	4	39	5
30	-1	-1	-1	-1	-1	3	3	2	4	4	35	9
31	-1	-1	-1	-1	-1	3	3	2	4	4	52	8
32	-1	-1	-1	-1	-1	3	3	2	4	4	42	2

A.2.2 DCE for Asthma Health States Using Bayesian Approach

A.2.2.1 Bayesian Designs with Full Profiles

Table 2: Bayesian paired comparisons with full profiles for asthma health states generated based on the $\mathcal{D}^{B}_{\mathbf{s},\mathrm{FIM}}$ and assuming Gamma(1,10) prior distribution for the logit model parameters

Choice Set	**Alternative I**					**Alternative II**				
1	-1	-1	-1	-1	-1	4	4	4	4	4
2	3	2	1	4	1	4	1	2	2	3
3	0	4	1	2	4	1	3	2	3	1
4	4	3	3	2	1	3	2	4	1	2
5	3	4	2	3	2	2	3	3	4	3
6	1	2	4	2	3	3	1	3	4	2
7	2	1	0	1	4	3	0	1	2	3
8	4	4	2	0	3	2	3	3	2	2
9	1	1	1	4	4	2	3	0	3	3
10	0	2	3	1	4	2	1	4	2	3
11	4	0	3	3	1	3	1	2	2	2
12	1	0	0	1	2	2	2	1	0	1
13	1	1	1	2	3	2	0	2	1	4
14	0	4	1	1	3	3	2	0	0	4
15	1	3	2	4	3	4	4	1	3	2
16	0	4	4	4	1	1	3	2	3	2
17	2	2	2	3	2	3	3	0	1	4
18	1	4	3	2	2	0	1	4	4	3
19	1	1	3	3	2	0	3	2	4	1
20	4	3	1	0	2	3	4	0	2	1
21	0	0	1	2	1	1	2	3	1	0
22	2	0	1	2	2	0	1	2	3	1
23	1	0	2	0	4	3	1	0	1	3
24	2	1	1	3	4	3	0	3	1	3
25	0	2	4	2	1	3	1	1	0	3
26	0	2	3	2	3	1	3	4	1	2
27	2	4	1	3	3	4	1	4	2	2
28	3	2	1	2	4	2	4	3	0	2
29	3	2	3	3	3	4	4	0	4	1
30	4	2	2	2	2	1	0	4	3	4
31	2	0	1	1	0	1	3	0	0	2
32	3	3	4	2	1	4	2	0	4	2

Table 3: Bayesian paired comparisons with full profiles for asthma health states generated based on the $\mathcal{D}^{B}_{\mathbf{s},\text{GFIM}}$ and assuming Gamma(1,10) prior distribution for the logit model parameters

Choice Set	**Alternative I**					**Alternative II**				
1	-1	-1	-1	-1	-1	4	4	4	4	4
2	0	1	3	1	2	1	0	2	0	4
3	4	1	2	0	1	3	0	1	2	2
4	4	3	2	3	2	2	4	4	2	1
5	3	4	2	2	4	2	3	4	3	3
6	3	0	3	4	4	4	2	4	1	3
7	0	2	4	0	4	1	1	2	2	3
8	1	2	3	4	3	4	1	4	2	2
9	2	3	0	2	0	0	4	1	1	1
10	4	4	3	2	1	2	1	2	4	2
11	2	1	2	3	3	3	3	3	1	1
12	2	4	1	1	4	0	3	3	2	2
13	0	1	1	3	3	1	0	4	1	2
14	4	4	1	0	2	3	1	2	1	4
15	2	1	1	4	0	1	2	0	3	1
16	4	0	3	4	1	3	3	1	3	2
17	4	2	3	0	2	0	1	4	2	3
18	3	4	3	3	2	4	3	1	4	3
19	1	2	1	4	0	2	3	0	2	3
20	1	2	1	4	2	0	4	2	1	1
21	4	2	0	3	1	1	3	3	1	3
22	0	0	2	1	0	2	1	0	0	1
23	1	0	4	0	4	2	4	0	1	1
24	2	2	0	4	2	1	3	1	3	1
25	3	1	3	0	3	2	3	2	2	1
26	3	2	1	2	2	1	4	0	4	1
27	3	2	2	4	3	2	1	3	3	4
28	2	2	3	2	4	3	0	4	3	3
29	1	4	1	2	3	3	3	2	3	1
30	1	1	4	4	2	2	0	3	3	3
31	1	2	3	0	0	0	0	0	1	2
32	2	2	2	1	3	0	3	0	4	4

A.2.2.2 Bayesian Designs with Partial Profiles

Table 4: Bayesian paired comparisons with partial profiles for asthma health states generated based on the $\mathcal{D}^{B}_{\mathbf{s},\text{FIM}}$ and assuming Gamma(1,10) prior distribution for the logit model parameters

Choice Set	**Alternative I**					**Alternative II**				
1	-1	-1	-1	-1	-1	4	4	4	4	4
2	*	*	3	4	1	*	*	4	3	0
3	*	*	3	1	0	*	*	2	0	2
4	*	*	3	2	4	*	*	4	4	3
5	*	0	*	0	4	*	1	*	2	2
6	*	4	*	2	3	*	3	*	3	4
7	*	0	*	4	3	*	4	*	3	1
8	*	0	1	*	2	*	1	3	*	0
9	*	0	3	*	1	*	1	0	*	4
10	*	3	2	*	3	*	4	1	*	4
11	*	1	4	*	1	*	4	2	*	0
12	*	3	2	1	*	*	2	1	3	*
13	*	2	4	2	*	*	1	3	3	*
14	*	0	2	2	*	*	3	1	0	*
15	3	*	*	4	0	2	*	*	3	3
16	2	*	*	4	4	3	*	*	3	3
17	0	*	*	2	3	1	*	*	1	2
18	4	*	1	*	3	3	*	4	*	1
19	3	*	3	*	2	4	*	2	*	1
20	0	*	3	*	2	2	*	1	*	1
21	1	*	1	1	*	2	*	0	0	*
22	1	*	4	2	*	3	*	2	3	*
23	4	*	3	0	*	2	*	0	4	*
24	1	1	*	*	4	2	3	*	*	2
25	0	2	*	*	1	4	0	*	*	0
26	4	2	*	*	2	3	3	*	*	3
27	0	4	*	3	*	3	2	*	2	*
28	1	3	*	2	*	4	2	*	1	*
29	0	4	*	1	*	1	3	*	3	*
30	2	1	1	*	*	1	2	0	*	*
31	0	3	1	*	*	1	4	0	*	*
32	2	2	2	*	*	4	3	0	*	*

Table 5: Bayesian paired comparisons with partial profiles for asthma health states generated based on the $\mathcal{D}^{B}_{\mathbf{S},\text{GFIM}}$ and assuming Gamma(1,10) prior distribution for the logit model parameters

Choice Set	**Alternative I**					**Alternative II**				
1	-1	-1	-1	-1	-1	4	4	4	4	4
2	*	*	1	4	1	*	*	4	2	0
3	*	*	0	4	3	*	*	2	0	4
4	*	*	3	2	4	*	*	2	3	3
5	*	0	*	1	4	*	1	*	2	2
6	*	0	*	4	3	*	2	*	3	2
7	*	2	*	4	2	*	4	*	3	1
8	*	0	2	*	2	*	1	3	*	0
9	*	0	3	*	1	*	1	0	*	4
10	*	3	2	*	3	*	4	0	*	2
11	*	1	4	*	1	*	4	2	*	0
12	*	1	2	4	*	*	3	1	3	*
13	*	2	4	2	*	*	1	3	3	*
14	*	0	2	3	*	*	3	1	0	*
15	4	*	*	4	0	3	*	*	2	3
16	0	*	*	4	4	3	*	*	1	3
17	0	*	*	0	3	1	*	*	1	2
18	1	*	3	*	3	0	*	4	*	2
19	0	*	3	*	2	4	*	0	*	1
20	0	*	0	*	3	3	*	1	*	0
21	0	*	1	1	*	2	*	0	0	*
22	2	*	4	3	*	4	*	2	2	*
23	1	*	4	0	*	2	*	1	2	*
24	1	4	*	*	2	2	3	*	*	1
25	1	2	*	*	1	4	0	*	*	2
26	4	2	*	*	3	3	1	*	*	4
27	3	2	*	3	*	2	3	*	2	*
28	3	0	*	2	*	2	2	*	1	*
29	0	4	*	1	*	1	0	*	3	*
30	4	1	1	*	*	3	3	0	*	*
31	1	3	2	*	*	3	2	0	*	*
32	2	4	1	*	*	4	3	0	*	*

Table 6: Bayesian paired comparisons with partial profiles for asthma health states generated based on the $\mathcal{D}^{B}_{\mathbf{S},\text{FIM}}$ and assuming TTO prior distribution for the logit model parameters

Choice Set	**Alternative I**					**Alternative II**				
1	-1	-1	-1	-1	-1	4	4	4	4	4
2	*	*	1	3	0	*	*	3	0	1
3	*	*	0	2	2	*	*	4	0	0
4	*	*	2	3	1	*	*	3	4	0
5	*	0	*	3	1	*	2	*	1	2
6	*	2	*	4	4	*	4	*	1	3
7	*	3	*	4	1	*	0	*	0	3
8	*	3	2	*	0	*	1	4	*	1
9	*	0	3	*	4	*	3	0	*	3
10	*	2	4	*	4	*	3	3	*	3
11	*	4	4	*	0	*	3	1	*	2
12	*	0	4	1	*	*	2	3	3	*
13	*	2	2	4	*	*	4	3	0	*
14	*	3	4	2	*	*	4	1	4	*
15	0	*	*	4	2	1	*	*	1	4
16	4	*	*	3	2	2	*	*	0	3
17	1	*	*	2	1	4	*	*	0	2
18	1	*	4	*	2	2	*	0	*	4
19	0	*	1	*	3	2	*	2	*	2
20	4	*	2	*	1	3	*	0	*	0
21	0	*	4	3	*	4	*	0	2	*
22	3	*	2	2	*	2	*	0	3	*
23	2	*	1	2	*	3	*	0	1	*
24	0	3	*	*	4	4	2	*	*	3
25	4	1	*	*	0	3	0	*	*	2
26	3	4	*	*	1	0	1	*	*	3
27	2	0	*	1	*	1	1	*	0	*
28	0	4	*	2	*	1	3	*	1	*
29	0	2	*	1	*	1	0	*	4	*
30	0	0	3	*	*	3	1	1	*	*
31	2	1	2	*	*	4	2	1	*	*
32	1	4	0	*	*	3	1	3	*	*

Bibliography

Anderson, D. A. and Wiley, J. B. (1992). Efficient choice set designs for estimating availability cross-effects models, *Marketing Letters*, **3 (4)**: 357–370.

Ara, R. and Brazier, J. (2008). Deriving an algorithm to convert the eight mean SF-36 dimension scores into a mean EQ-5D preference-based score from published studies (where patient level data are not available), *Value in Health*, **11 (7)**: 1131–1143.

Arentze, T., Borgers, A., Timmermans, H. and DelMistro, R. (2003). Transport stated choice responses: effects of task complexity, presentation format and literacy, *Transportation Research Part E: Logistics and Transportation Review*, **39 (3)**: 229–244.

Atkinson, A. C., Donev, A. N. and Tobias, R. D. (2007). *Optimum experimental designs with SAS*, Oxford University Press.

Bansback, N., Brazier, J., Tsuchiya, A. and Anis, A. (2010). Using a discrete choice experiment to estimate societal health state utility values, Tech. rep., University Library of Munich, Germany.

Bansback, N., Hole, A. R., Mulhern, B. and Tsuchiya, A. (2014). Testing a discrete choice experiment including duration to value health states for large descriptive

systems: Addressing design and sampling issues, *Social Science & Medicine*, **114**: 38–48.

Barton, M. G. R., Bankart, J., Davis, A. C. and Summerfield, Q. A. (2004). Comparing utility scores before and after hearing-aid provision, *Applied Health Economics and Health Policy*, **3 (2)**: 103–105.

de Bekker-Grob, E. W., Ryan, M. and Gerard, K. (2010). Discrete choice experiments in health economics: A review of the literature, *Health Economics*, **(2)**: 145–172.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*, Springer.

Bleichrodt, H. (2002). A new explanation for difference between time-trade off utilities and standard gamble utilities, *Health Economics*, **11**: 447–456.

Bliemer, M. C., Rose, J. M. and Hess, S. (2008). Approximation of Bayesian efficiency in experimental choice designs, *Journal of Choice Modelling*, **1 (1)**: 98–126.

Brazier, J., Deverill, M., Green, C., Harper, R. and Booth, A. (1999). A review of the use of health status measures in economic evaluation, *Health Technology Assessment*, **3 (9)**.

Brazier, J., Roberts, J. and Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36, *Health Economics*, **21 (2)**: 271–292.

Brazier, J., Rowen, D., Mavranezouli, I., Tsuchiya, A., Young, T., Yang, Y., Barkham, M. and Ibbotson, R. (2012). Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome), *Health Technology Assessment*, **16 (32)**: 1366–5278.

Brazier, J., Rowen, D., Yang, Y. and Tsuchiya, A. (2009). Health economics and decision science discussion paper series no.09/10: Using rank and discrete choice data to estimaate health state utility values on the QALY scale, *Health Economics, University of Sheffield*.

Brazier, J. and Tsuchiya, A. (2010). Preference-based condition-specific measures of health: what happens to cross programme comparability?, *Health Economics*, **19 (2)**: 125–129.

Brazier, J. E., Ratcliffe, J., Tsuchiya, A. and Solomon, J. (2007). *Measuring and valuing health benefits for economic evaluation*, Oxford University Press.

Brazier, J., Y. Y. and Tsuchiya, A. (2006). Using rank and discrete choice data to estimate health state utility values for QALYs: the case of the AQL-5D, *Health Economics study group meeting*.

Brooks, R. (1996). Euroqol: the current state of play, *Health Policy*, **37 (1)**: 53–72.

Bryan, S., Buxton, M., Sheldon, R. and Grant, A. (1998). Magnetic resonance imaging for the investigation of knee injuries: An investigation of preference, *Health Economics*, **7**: 595604.

Bunch, D. S., Louviere, J. J. and Anderson, D. (1996). A comparison of experimental design strategies for multinomial logit models: The case of generic attributes, *Graduate School of Management, University of California at Davis*.

Burr, J. M., Kilonzo, M., Vale, L. and Ryan, M. (2007). Developing a preference-based glaucoma utility index using a discrete choice experiment, *Optometry & Vision Science*, **84 (8)**: E797–E809.

Cain, T. (2011). *Bayesian inference for health state utilities using pairwise comparsion data*, Thesis, Sheffield University.

Carlsson, F. and Martinsson, P. (2003). Design techniques for stated preference methods in health economics, *Health Economics*, **12 (4)**: 281–294.

Carson, R. T., Louviere, J. J., Anderson, D. A., Arabie, P., Bunch, D. S., Hensher, D. A., Jonson, R. M., Kuhfeld, W. F., Steinberg, D., Swait, J., Timmermans, H. and Wiley, J. B. (1994). Experimental analysis of choice, *Marketing Letters*, **5 (351-367)**.

Caussade, S., Ortúzar, J. d. D., Rizzi, L. I. and Hensher, D. A. (2005). Assessing the influence of design dimensions on stated choice experiment estimates, *Transportation research part B: Methodological*, **39 (7)**: 621–640.

Chaloner, K. (1984). Optimal experimental design for linear models, *The Annals of Statistics*, **12 (1)**: 283–300.

Chaloner, K. and Verdinelli, I. (1995). Bayesain experimental design: A review, *Statistical Science*, **10 (3)**: 273–304.

Clark, M. D., Determann, D., Petrou, S., Moro, D. and de Bekker-Grob, E. W. (2014). Discrete choice experiments in health economics: A review of the literature, *Pharmaco Economics*, pp. 1–20.

Cook, R. D. and Nachtsheim, C. J. (1980). A comparison of algorithm for constructing exact $\mathcal{D}$-optimal designs, *Technometrics*, **22**: 315–324.

Cuervo, D. P., Kessels, R., Goos, P. and Sörensen, K. (2015). An integrated algorithm for the optimal design of stated choice experiments with partial profiles, *UNIVERSITY OF ANTWERP, Faculty of Applied Economics*.

Dolan, P., Gudex, C., Kind, P. and Williams, A. (1996a). The time trade-off method: Results from a general population study, *Health Economics*, **5 (2)**: 141–154.

Drummond, M. F., Sculpher, M. J., Torrance, G. W., O'Brien, B. J. and Stoddart, G. L. (2005). *Methods for the economic evaluation of health care programmes*, Oxford University Press, third edn.

Dykstra, O. (1971). The augmentation of experimental data to maximize $|\mathbf{X}^T\mathbf{X}|$, *Technometrics*, **13 (3)**: 682–688.

Earnshaw, J. and Lewis, G. (2008). Nice guide to the methods of technology appraisal, *Pharmaco Economics*, **26 (9)**: 725–727.

Espallargues, M., Czoski-Murray, C. J., Bansback, N. J., Carlton, J., Lewis, G. M., Hughes, L. A., Brand, C. S. and Brazier, J. E. (2005). The impact of age-related macular degeneration on health status utility values, *Investigative Ophthalmology & Visual Science*, **46 (11)**: 4016–4023.

EuroQol, G. (1990). Euroqol-a new facility for the measurement of health-related quality of life, *Health Policy (Amsterdam, Netherlands)*, **16 (3)**: 199.

Federov, V. V. (1972). *Theory of optimal experiments*, New York Academic Press.

Flynn, T. (2009). Using conjoint analysis and choice experiments to estimate QALY values: issues to consider, *Pharmaco Economics*, **28 (9)**: 711–722.

Flynn, T. N. (2010). Using conjoint analysis and choice experiments to estimate QALY values, *Pharmaco Economics*, **28 (9)**: 711–722.

Franks, P., Lubetkin, E. I., Gold, M. R., Tancredi, D. J. and Jia, H. (2004). Mapping the SF-12 to the euroqol EQ-5D index in a national us sample, *Medical Decision Making*, **24 (3)**: 247–254.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004). *Bayesian data analysis*, Chapman & Hall/CRC, second edn.

Goos, P., Vermeulen, B. and Vandebroek, M. (2010). D-optimal conjoint choice designs with no-choice options for a nested logit model, *Journal of Statistical Planning and Inference*, **140 (4)**: 851–861.

Gray, A. M., Rivero-Arias, O. and Clarke, P. M. (2006). Estimating the association between SF-12 responses and EQ-5D utility values by response mapping, *Medical Decision Making*, **26 (1)**: 18–29.

Green, P. E. (1974). On the design of choice experiments involving multifactor alternatives, *Journal of Consumer Research*, pp. 61–68.

Gu, N. Y., Botteman, M. F., Gerber, R. A., Ji, X., Postema, R., Wan, Y., Sianos, G., Anthony, I., Cappelleri, J. C., Szczypa, P. et al. (2013). Eliciting health state utilities for dupuytren's contracture using a discrete choice experiment, *Acta Orthopaedica*, **84 (6)**: 571–578.

Gudex, C. (1994). Time trade-off user manual: props and self-completion methods, Tech. rep., Centre for Health Economics, University of York.

Huber, J. and Zwerina, K. (1996). The importance of utility balance in efficient choice designs, *Journal of Marketing Research*, **33 (3)**: 307–317.

Johnson, F. R., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D., Bresnahan, B. W., Kanninen, B. and Bridges, J. F. P. (2013). Constructing experimental designs for discrete-choice experiments: report of the ispor conjoint analysis experimental design good research practices task force, *Value in Health*, **16 (1)**: 3–13.

Johnson, R. M. and Orme, B. K. (1996). How many questions should you ask in choice-based conjoint studies, *Sawtooth software technical paper*.

Kessels, R. (2010). Design and analysis of choice experiments in JMP, Antwerpen University.

Kessels, R., Goos, P. and Vandebroek, M. (2004). Comparing algorithms and criteria for designing Bayesian conjoint choice experiment, *DTEW Research Report 0427*, pp. 1–38.

Kessels, R., Goos, P. and Vandebroek, M. (2006). A comparison of criteria to design efficient choice experiments, *Journal of Marketing Research*, **43**: 409–419.

Kessels, R., Goos, P. and Vandebroek, M. (2008). Optimal designs for conjoint choice experiments, *Computational Statistics and Data Analysis*, **52**: 2369–2387.

Kessels, R., Jones, B. and Goos, P. (2011a). Bayesian optimal designs for discrete choice experiments with partial profiles, *Journal of Choice Modelling*, **4 (3)**: 52–74.

Kessels, R., Jones, B. and Goos, P. (2014). An improved two-stage variance balance approach for constructing partial profile designs for discrete choice experiments, *Applied Stochastic Models in Business and Industry*.

Kessels, R., Jones, B., Goos, P. and Vandebroek, M. (2009). An efficient algorithm for constructing Bayesian optimal choice designs, *Journal of Business and Economic Statistics*, **27 (2)**: 279–291.

Kessels, R., Jones, B., Goos, P. and Vandebroek, M. (2011b). The usefulness of Bayesian optimal designs for discrete choice experiments, *Applied Stochastic Models in Business and Industry*, **27 (3)**: 197–203.

Kiefer, J. and Wolfowitz, J. (1959). Optimum designs in regression problems, *The Annals of Mathematical Statistics*, pp. 271–294.

Kiefer, J. and Wolfowitz, J. (1960). The equivalence of two extremum problems, *Canadian Journal of Mathematics*, **12**: 363–366.

Kuhfeld, W. (2000). *Marketing research methods in the SAS system*, SAS Institute: Nary, NC, USA, 8th edn.

Kuhfeld, W. F., Tobias, R. D. and Garratt, M. (1994). Efficient experimental design with marketing research applications, *Journal of Marketing Research*, **12**: 545–557.

Lindley, D. V. (1956). On a measure of the information provided by an experiment, *The Annals of Mathematical Statistics*, pp. 986–1005.

Louviere, J. J., Hensher, D. A. and Swait, J. D. (2000). *Stated choice methods: analysis and applications*, Cambridge University Press.

Lunn, D.J., T. A. B.-N. and Spiegelhalter, D. (2000). Winbugs a Bayesian modelling framework: concepts, structure, and extensibility, *Statistics and Computing*, **10**: 325–337.

Marschak, J. (1960). Binary choice constraints on random utility indications, *Stanford Symposium on Mathematical Methods in the Social Sciences, Stanford University Press*, pp. 312–329.

Marshall, D., Bridges, J. F., Hauber, B., Cameron, R., Donnalley, L., Fyie, K. and Johnson, F. R. (2010). Conjoint analysis applications in healthhow are studies being designed and reported?, *The Patient: Patient-Centered Outcomes Research*, **3 (4)**: 249–256.

McCabe, C., Brazier, J., Gilk, P., Tsuchiya, A., Roberts, J., O'Hagan, A. and Stevens, K. (2006). Using rank data to estimate health state utility models, *Health Economics*, **25**: 418–431.

McCulloch, R. and Rossi, P. E. (1994). An exact likelihood analysis of the multinomial probit model, *Econometrics*, **64 (1)**: 207–240.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour, *Frontiers in Econometrics*, pp. 105–142.

Meyer, R. K. and Nachtsheim, C. J. (1995). The coordinate-exchange algorithm for constructing exact optimal experiment designs, *Technometrics*, pp. 60–69.

Mitchell, T. J. (1974). An algorithm for constructing of $\mathcal{D}$-optimal experimental designs, *Technometrics*, **16 (2)**: 203–210.

O'Hagan, A., Forster, J. and Kendall, M. G. (2004). *Bayesian inference*, Arnold London.

Pol, M. V. D. and Cairns, J. (1998). Establishing preferences for blood transfusion support: An application of conjoint analysis, *Health Services Research and Management*, **3**: 70–76.

Propper, C. (1990). Contingent valuation of time spent on NHS waiting list, *The Economic Journal*, **100**: 193–199.

Prosser, L. A., Payne, K., Rusinak, D., Shi, P. and Messonnier, M. (2013). Using a discrete choice experiment to elicit time trade-off and willingness-to-pay amounts for influenza health-related quality of life at different ages, *Pharmaco Economics*, **31 (4)**: 305–315.

R Core Team (2013). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.

Rady, E. A., El-Monsef, M. M. E. A. and Seyam, M. M. (2009). Relationships among several optimality criteria, *InterStat*, **247**: 1–11.

Ramos-Goñi, J. M., Rivero-Arias, O., Errea, M., Stolk, E. A., Herdman, M. and Cabasés, J. M. (2013). Dealing with the health state dead when using discrete choice experiments to obtain values for EQ-5D-5l heath states, *The European Journal of Health Economics*, **14 (1)**: 33–42.

Ratcliffe, J., Brazier, J., Tsuchiya, A., Symonds, T. and Brown, M. (2009). Using DCE and ranking data to estimate cardinal values for health states for deriving a preference-based single index from the sexual quality of life questionnaire, *Health Economics*, **18 (11)**: 1261–1276.

Rose, J. and Bliemer, C. (2008). *Stated preference experimental design strategies*, Elsevier.

Rose, J. and Bliemer, C. (2012). Ngene, Available from: http://www.choice-metrics.com/download.html. [Accessed July 21, 2013].

Rose, J. M. and Bliemer, M. C. (2009). Constructing efficient stated choice experimental designs, *Transport Reviews*, **29 (5)**: 587–617.

Rose, J. M. and Bliemer, M. C. (2013). Sample size requirements for stated choice experiments, *Transportation*, **40 (5)**: 1021–1041.

Rose, J. M., Bliemer, M. C., Hensher, D. A. and Collins, A. T. (2008). Designing efficient stated choice experiments in the presence of reference alternatives, *Transportation Research Part B: Methodological*, **42 (4)**: 395–406.

Rowen, D., Brazier, J., Hout, B. et al. (2011). A comparison of methods for converting dce values onto the full health-dead QALY scale, *HEDS Discussion Paper 11/15*.

Ryan, M. (1999). A role for conjoint analysis in technology assessment in health care, *Technology Assessment in Health Care*, **15 (3)**: 433–457.

Ryan, M. and Farrar, S. (1994). A pilot study using conjoint analysis to establish the views of users in the provision of orthodontic services in Grampian. Health Economics Research Unit discussion paper no. 07/94, *Aberdeen: University of Aberdeen*.

Ryan, M. and Gerard, K. (2003). Using discrete choice experiment to value health care programmes: Current practice and future research reflection, *Applied Health Economics and Health Policy*, **1**: 55–64.

Ryan, M. and Hughes, J. (1997). Using conjoint analysis to value surgical versus medical management of miscarriage, *Health Economics*, **6**: 26173.

Ryan, M., Netten, A., Skatun, D. and Smith, P. (2006). Using discrete choice experiments to estimate a preference-based measure of outcome-an application to social care for older people, *Journal of Health Economic*, **25**: 927–944.

Sándor, Z. and Wedel, M. (2001). Designing conjoint choice experiments using managers' prior beliefs, *Journal of Marketing and Research*, **38 (4)**: 430–444.

Sándor, Z. and Wedel, M. (2002). Profile construction in experimental choice designs for mixed logit models, *Marketing Science*, **21 (4)**: 455–475.

Schwabe, R., Grasshoff, U., Grossmann, H. and Holling, H. (2003). Optimal 2^k paired comparison designs for partial profiles, *Tatra Mountains Mathematical Publication*, **26**: 79–86.

Shannon, C. E. (1948). A mathmatical theory of communication, *Bell System Technical Journal*, **27**: 379–423.

Silvey, S. D. (1980). *Optimal design*, vol. 7, Springer.

Smith, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations, *Biometrika*, pp. 1–85.

Stolk, E. A., Oppe, M., Scalone, L. and Krabbe, P. F. (2010). Discrete choice modeling for the quantification of health states: The case of the EQ-5D, *Value in Health*, **13 (8)**: 1005–1013.

Street, D. J. and Burgess, L. (2007). *The construction of optimal stated choice experiments: Theory and methods*, vol. 647, John Wiley and Sons.

Street, D. J., Burgess, L. and Louviere, J. J. (2005). Quick and easy choice sets: constructing optimal and nearly optimal stated choice experiments, *Research in Marketing*, **22 (4)**: 459–470.

Sturtz, S., Ligges, U. and Gelman, A. E. (2005). R2WinBUGS: a package for running WinBUGS from R, *Statistical software*, **12 (3)**: 1–16.

Thurstone, L. (1927). A law of comparative judgement, *Psychological Review*, **34**: 273–286.

Tolley, K. (2009). What are health utilities, *Hayward Medical Communications, London*.

Torrance, G. W., Thomas, W. H. and Sackett, D. L. (1972). A utility maximization model for evaluation of health care programs, *Health Services Research*, **7 (2)**: 118.

Train, K. (2003). *Discrete choice methods with simulation*, Cambridge University Press, second edn.

Viney, R., Norman, R., Brazier, J., Cronin, P., King, M. T., Ratcliffe, J. and Street, D. (2014). An australian discrete choice experiment to value EQ-5D health states, *Health Economics*, **23 (6)**: 729–742.

Wald, A. (1943). On the efficient design of statistical investigations, *The Annals of Mathematical Statistics*, **14 (2)**: 134–140.

Walters, S. J., Morrell, C. J. and Dixon, S. (1999). Measuring health-related quality of life in patients with venous leg ulcers, *Quality of Life Research*, **8 (4)**: 327–336.

Weinstein, M., Siegel, J., Gold, M., Kamlet, M. and Russell, L. (1996). Cost-effectiveness in health and medicine, *New York: Oxford University*.

Yang, Y., Brazier, J. E., Tsuchiya, A. and Young, T. A. (2007). Estimating a preference-based single index from the asthma qality of life questionnaire (AQLQ), *MPRA*.

Yang, Y., Brazier, J. E., Tsuchiya, A. and Young, T. A. (2010). Estimating a preference-based index for a 5-dimensional health state classification for asthma derived from the asthma quality of life questionnaire, *Medical Decision Making*, p. 0272989X10379646.

Yu, J., Goos, P. and Vandebroek, M. (2008). A comparison of different Bayesian design criteria to compute efficient conjoint choice experiments, *Available at SSRN 1288549*.

Zwerina, K., Huber, J. and Kuhfeld, W. F. (1996). A general method for constructing efficient choice designs, *Statistical Research and Development, SAS Institute*, pp. 265–282.