Università
Ca'Foscari
Venezia

Corso di Dottorato di ricerca

in Economia

ciclo 31

Tesi di Ricerca

STATED PREFERENCES IN HEALTH ECONOMICS

SSD: SECS-P01

Coordinatore del Dottorato
ch. prof. Giacomo Pasini

Supervisore
ch. prof. Giacomo Pasini

Supervisore cotutela
ch. prof. Antonio Nicolò

Dottorando
Mesfin Geremew
Genie
Matricola 956252

Università Ca'Foscari Venezia

Ca' Foscari University of Venice

Department of Economics
Corso di Dottorato in Economia

Cycle 31

Stated Preferences in Health Economics

Candidate
Mesfin Geremew Genie

Supervisors
Prof. Giacomo Pasini

Prof. Antonio Nicolò

PhD Coordinator
Prof. Giacomo Pasini

Cycle 31, 2015-2018

PhD Thesis Committee:

Prof. Daniel Rigby, The University of Manchester

Prof. Jürgen Maurero, Université de Lausanne

Prof. Luca Corazzini, Ca' Foscari University of Venice

Declaration of Authorship

I, Mesfin GENIE, declare that this thesis titled, ”Stated Preferences in Health Economics” and the work presented in it are my own, although some of the chapters are based upon permitted collaborative research. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- No part of the thesis has been submitted in support of an application for another degree or any other university.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given.
- I have acknowledged all main sources of help.

Signed:

Genie

Date: 12/12/2018

For my unknown mother

Acknowledgments

I want to acknowledge and thank a lot of different people for their support. I am particularly grateful to Professor Giacomo Pasini and Professor Antonio Nicolò for the supervision, intellectual guidance, and patience throughout this PhD thesis preparation, without forgetting their good human relationships. Thank you for all the support and willingness to help. I gratefully acknowledge the funding received towards my PhD from the Ca' Foscari University of Venice. Many thanks also go out to the Health Economics Research Unit, the University of Aberdeen for hosting my visiting period and providing the necessary support in developing part of my thesis. I am especially grateful to Professor Mandy Ryan and Dr Nicolas Krucien, who has been a constant source of support concerning chapter 4 of the thesis. Many thanks also to Professor Luca Corazzini, who acted as an internal referee, for his constructive feedback on this thesis. I would also like to thank the external committee members: Professor Daniel Rigby (The University of Manchester) and Professor Jürgen Maurero (Université de Lausanne) for their time, interest, and helpful comments. I have tremendously benefited from the course on 'Discrete choice analysis: predicting demand and market share' at EPFL in Lausanne and I am grateful to Professor Ben-Akiva Moshe (MIT) and Professor Michel Bierlaire (EPFL, Lausanne) for providing a scholarship to attend the course. I want to thank Lisa Negrello for all the administrative related support and help provided in the Department of Economics. Finally, of course, the support from my friends has been invaluable. This is an excellent opportunity to express my thanks, respect and gratefulness for your friendship, encouragement and assistance.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Choice experiments (CEs) are commonly used in applied economics to value non-market goods, e.g., in health care, how do patients value and trade-off factors such as treatment effectiveness and risk of side effects in the delivery of healthcare. In a CE, respondents are typically asked to choose between two or more multi-attribute hypothetical descriptions of the good. These stated preferences are then used to estimate the marginal utility of changes in the composition of the good. This PhD thesis aims to contribute to the base of knowledge about stated preferences in some aspects of health economics. The thesis contains three independent papers presented as individual chapters, along with this introductory chapter and a concluding chapter that summarises the three studies and draws lessons from the research as a whole. In this introduction, I provide an overview of the research questions and discuss the context of each chapter. All the studies have in common that they feature a choice experiment. However, regarding content, various aspects of health economics are investigated. These studies have a novelty that a CE is applied to research questions that have not been investigated with this method before.

The first paper, '*Willingness to wait heterogeneity: Does it matter for kidney transplantation?*', investigated heterogeneity in patients' willingness to wait (WTW) for changes in the time and risk attributes of kidney transplantation and examined how the heterogeneity in WTW can be mapped with observable characteristics of the patients. Using mixed logit models in WTW-space, the paper provided evidence of heterogeneity in patients' WTW for changes in the attributes of kidney transplantation. We compare the entire distribution of WTW across different groups of patients and demonstrate that younger patients are willing to wait longer for an extra year of graft survival and to avoid augmented infectious and neoplastic risks. Moreover, patients with longer duration of dialysis are willing to wait longer for an extra year of graft survival and to avoid augmented infectious and neoplastic risks. The implication for transplant practice is that accounting for patient preferences in kidney allocation algorithm may improve patients' satisfaction and the donor-receiver matching process.

In the second paper, '*Does cognitive ability affect choice consistency?*', the effect of patients' cognitive ability on the consistency of responses obtained from CEs are explored using heteroskedastic multinomial logit (HMNL), generalised multinomial logit (GMNL) models, and the same data set as in the first paper. Both the HMNL and GMNL models indicated that patients' cognitive ability did impact significantly on the consistency of the CE responses. A higher cognitive ability (measured by numeracy score) tended to result in a higher scale (a lower variance of the error term) thereby lower choice randomness and hence a consistent choice response in CEs. Patients who are consistent in their choices have a lower WTW for changes in the multi-attribute content of kidney transplantation. This study provided evidence that respondents' cognitive ability is an important determinant of choice consistency in CE responses.

In the third paper, '*Attributes aggregation in multi-attribute choices: Does it exist?*', the main objective is to investigate whether individuals aggregate multi-attribute information when completing choice tasks in CEs. A CE survey concerned with preferences for personalisation of chronic pain self-management programmes in the UK is used to explore attributes aggregation (AA) in multi-attribute choices. We develop a framework in which individuals restructure the multi-attribute information into a meta-attribute (e.g., convert non-monetary attributes into a single quality dimension) before making their decisions. We estimated a non-linear utility model allowing AA to depend on the information structure. This new model assumes participants are more likely to aggregate the quality information into a meta-attribute when the quality attributes provide similar information about the good or service. We find evidence of AA when responding to CEs, with the probability of adopting AA greater for homogenous information. AA is more prevalent amongst participants who adopted a quick and click strategy (shorter response time), more likely to occur for later positioned choice tasks (potentially due to fatigue effect), leads to improvements in model fit and has implications for welfare estimates. Our results underline the importance of accounting individuals' information processing rules when modelling multi-attribute choices.

1.1 The research questions

The aim of the whole doctoral thesis is to address three research questions in health economics:

1. Are transplant patients' WTW for changes in the time and risk attributes of kidney transplantation heterogeneous? Is it possible to map WTW to patients' observable characteristics?
2. Does cognitive ability (proxied by numeracy score) affect choice consistency in a choice experiment? Does consistency change the WTW estimates?
3. Do individuals aggregate multi-attribute information when completing choice tasks in choice experiments (CEs)? What are the implications of AA on the standard CE estimates such as WTP?

The first two questions are based on a stated preference survey concerned with preferences for time and risk attributes of kidney transplantation in Italy. The third question is addressed based on a CE survey concerned with preferences for personalisation of chronic pain self-management programmes in the UK.

1.2 Structure of this thesis

The thesis is organised in five chapters, and all the studies have in common that they feature a choice experiment. The first chapter is an introduction to the whole thesis. The second and third chapters are based on different parts of the same data set and are concerned mainly with the time and risk preferences in kidney transplantation. The fourth is based on another data set and focuses on the preferences for personalisation of chronic pain self-management programmes in the UK. The last chapter concludes the whole thesis.

Chapter 2

Willingness to wait heterogeneity: Does it matter for kidney transplantation?

ABSTRACT

Kidney transplantation provides an expected survival advantage over dialysis treatment for patients with the end-stage renal disease. However, due to the disparity between a large number of transplant candidates and scarcity of organs, patients may face the trade-off between a long waiting list for a high-quality kidney, or a "marginal" organ transplanted immediately. Current allocation protocols do not explicitly take into account patients' preferences. A shift towards implementing patient-centred care requires better insight into their preferences. We study patients' time and risk preferences for kidney transplantation in Italy using a choice experiment (CE). Using mixed logit models in WTW-space, we find heterogeneity in the patients' WTW for changes in the attributes of kidney transplantation. Differences are not limited to mean estimates: we compare the entire distribution of WTW across different groups of patients and demonstrate that younger patients are willing to wait longer for an extra year of graft survival and to avoid augmented infectious and neoplastic risks. Moreover, patients with longer duration of dialysis are willing to wait longer for a kidney that will offer an extra year of survival and to avoid augmented infectious and neoplastic risks. The implication for transplant practice is that accounting for patients' preferences in kidney allocation algorithm may improve patients' satisfaction and the donor-receiver matching process. [1] [2]

[1]This chapter is a joint work with Giacomo Pasini (Ca' Foscari University of Venice) and Antonio Nicolò (University of Padua).

[2]*Acknowledgments:* The authors wish to thank Noemi Pace, Lorenzo Rocco, Lucrezia Furian, Paolo Rigotti for helpful comments, Giacomo Battiston and Veronica Buizza for excellent assistance in the implementation of the experiment. Antonio Nicolò and Giacomo Pasini acknowledge the financial support of the Progetto di Ateneo KIDNEY from University of Padua. The authors are grateful for their comments to Mandy Ryan and participants of the HESG-Aberdeen, EuHEA-Lausanne, EEEA-ESEM Lisbon, iHEA 2017 Boston, seminar participants Padova, VIVE-Copenaghen (DKDK) as well as Dr. Sebastian Heidenreich and other participants of HERU stated preference seminar at the University of Aberdeen.

2.1 Introduction

Kidney transplantation carries several advantages over dialysis treatment for patients with end-stage renal disease (ESRD) in terms of long-term mortality risk, improved survival advantages and quality of life (Merion et al. 2005; Salvioli et al. 2016). Nevertheless, the disparity between a large number of transplant candidates and the scarcity of organs available continues to increase (Courtney and Maxwell, 2008); forcing patients to long waiting time, and stimulating transplant physicians to push the limits of donor suitability to utilise organs from donors with characteristics different from the "ideal" situation (so-called Expanded Criteria Donors (ECD))[3]. In this setting, the selection criteria for donor appropriateness have been widened significantly in recent years, including older persons and those with co-morbidities such as hypertension, diabetes, suboptimal renal function, or risky behaviours which may potentially increase the risk of infectious disease transmission.

An increasing number of transplants are now performed by expanding the pool of donors including those who would have been considered unsuitable before. The ECD program implemented since 2002 in the US and the Eurotransplant Seniors Program (ESP) implemented since 1999 in Europe are two examples of such policies. For instance, ECD or "marginal" kidneys, while inferior to standard criteria donor (SCD) kidneys, may prolong the life of the recipient compared to dialysis[4] treatment. Apart from survival advantage, an economic analysis also suggested that transplantation with a marginal donor kidney is more cost-effective than dialysis treatment (Whiting et al., 1999). The result of kidney transplantation from such marginal donors is one of the hottest topics in the transplant literature (Ojo et al. 2001; Metzger et al. 2003; Merion et al. 2005).

With the latest presumption, many transplant centres refuse to utilise kidney from marginal donors; therefore a significant number of kidneys are currently discarded. Unfortunately, there are no reliable and unambiguous means to define the outcome of transplantation for an organ. Several aspects related both to the donor quality and the recipient clinical conditions may affect the functional recovery, as well as the length of the cold ischemia time, defined as the interval between the procurement of the organ and its reperfusion during the recipient operation. Because kidneys start to degrade during this cold ischemic time, surgeons typically hope to transplant them within 24 hours. It has been claimed that

[3]ECD are deceased donor kidneys conveying a 70% or higher risk for a graft loss for transplant recipients relative to the ideal donation and are characterised by a donor age older than 60 years or older than 50 years and accompanied by two additional risk factors, including a history of hypertension, elevated terminal donor creatinine, and cerebrovascular cause of death (Metzger et al., 2003)).

[4]Dialysis is the process whereby blood is cycled out of the body and filtered through a machine, which removes waste and excess fluids.

organs discarded could be transplanted if the system for allocating them better matched the right organ to the proper recipient in the right amount of time. Sometimes, kidneys are discarded because the allocation process has required a too long time, for example when an organ is offered to several centres who refuse it (either the physicians or the patients) so that finally it becomes unsuitable.

Kidney recipients have a very important frontline role in defining how organs are allocated, and yet their preferences have been largely ignored in kidney allocation algorithms. In the transplantation of other organs such as the liver, heart, and lung, outside options are considerably limited. Vice versa, dialysis maintenance could be a reasonable option against which patients on the waiting list can balance risks and preferences. As a result, different patients may have heterogeneous preferences regarding the proposed treatment, i.e. regarding quality and waiting time. They may prefer to wait for either a long time with the prospect of receiving an "ideal" kidney or accept an organ of inferior quality with the advantage of short waiting time. Preferences may or may not correlate with social, cultural, economic status and psychological predispositions.

Despite the unique features of kidney transplantation, potential heterogeneity in patients' preferences has been largely ignored in organ allocation algorithms. Patients are informed regarding the risk factors of the donor they will receive the organ from, but at the time of entering the waiting list, they hardly have the chance to express their preferences towards the quality of organs they are willing to accept. The decision depends solely on medical/immunological compatibility, and it is made somewhat "automatic" by the allocation algorithm with limited involvement from the patients. One of the reasons why allocation algorithms do not account for patients' preferences is that pinning down preferences in a consistent way within the pool of transplant candidates is not an easy task. ESRD patients are unlikely to have the possibility to choose the medical treatment they have to go through; therefore it is not possible to infer their preferences from actual choices. We employed a stated preferences experiment to overcome this problem.

A limited but growing number of stated preference experiments have been conducted to address the general public and patients' preferences for various aspects of kidney transplantation. See Clark et al. (2018) for a systematic review of discrete choice experiments and conjoint analysis studies measuring trade-offs in nephrology. Howard et al. (2015) examined community respondents' preferences for the allocation of donor organs for transplantation (including kidneys and other organs) in Australia. The study suggested that allocation to younger patients were preferred over older patients. Family member donor registration, having caring responsibilities, and longer time on waiting list increased the priority. The same authors investigated community respondents' preferences for organ donation policy in Australia (Howard et al., 2016). The result showed a strong preference for a new organ donation policy that involved an easy registration process, an involvement of the donor's family in the final decision, direct payment or funeral expense reimbursement, and formal recognition of donation.

Reese et al. (2010) assessed patients' acceptability of a kidney from a donor with an increased risk of blood-borne viral infection (DIRVI) in the USA. The result indicated that patients were more likely to accept such kidneys the longer waiting time, the younger donor age, the lower HIV risk. Patients on dialysis and older participants were more likely to receive DIRVI kidneys. Clark et al. (2009) explored patients' preferences for a kidney transplant allocation criteria in the UK. The result suggested a strong preference for prioritising patients with moderate (such as uncontrolled hypertension or obesity plus kidney disease), not severe (such as heart attack, diabetes with complications, or stroke), diseases affecting life expectancy, improvement in kidney survival, having an extra-dependent adult or child, having no condition other than kidney disease affecting the quality of life (QoL), and having moderate rather than severe diseases affecting QoL. Clark et al. (2012) employed the same DCE to 908 patients, 41 carers, 113 healthcare professionals and 48 live donors or relatives of deceased donors. The study indicated a strong preference for prioritising patients with moderate diseases affecting life expectancy, a 1% improvement in survival of the kidney, having an extra adult or child, a one-year reduction in patient age, having no disease other than the kidney disease affecting QoL, and having moderate rather than severe diseases affecting QoL. The preferences of healthcare professionals' differed from those of patients for 5/7 variables. Small sample sizes limited assessment of preferences for live donors or relatives of deceased donors and carers.

Kamran et al. (2017) evaluated patients' preferences for accepting or not a marginal graft, for being informed about this type of graft, and for being involved in the decision-making process. Patients registered on the waiting list or already transplanted in eight transplant teams covering four main organs (i.e., kidney, liver, heart, and lung) in France participated in the experiment. The authors showed that 89% of patients were ready to accept a marginal graft, 76% preferred to be informed about these grafts but only 43% preferred to be involved in the decision-making process. They indicated that a marginal graft could be more accepted by patients who are in a critical medical situation or who perceive their condition as critical. The authors proposed giving patients with a critical situation the choice between acceptance of a marginal graft with short waiting time and reduced mortality risk on the national waiting list and long waiting time for a standard graft that usually works best than a marginal graft but could be available too late or never.

Stated preference experiments were also used to address issues in dialysis maintenance studies, including nephrologists preferences for dialysis in elderly patients with end-stage kidney disease (ESKD) in Australia (Foote et al., 2014), patients willingness to switch dialysis modality from conventional to more frequent dialysis in the USA (Halpern et al., 2004), public preferences for the location of dialysis facilities in Greenland (Kjær et al., 2013), and preferences for dialysis modality among pre-dialysis patients and caregivers in Australia (Morton et al., 2012).

However, none of these studies has explicitly looked into heterogeneity in the patients' willingness to wait (WTW) for changes in time and risk attributes of kidney transplantation. Moreover, some of the previous studies focused on the general public preferences for organ allocation criteria and policy, ignoring the patients perspective. While the preferences of the general public may provide input in the design of allocation protocols, it may not truly reflect the patients' preferences. To the best of our knowledge, this study is the first to apply a choice experiment (CE) to investigate patients' preferences for the time and risk attributes of kidney transplantation and examine trade-offs for these attributes based on a WTW approach. It is suggested that, if the goal of the CE is to obtain WTW or willingness to pay (WTP) estimates, a direct estimation of WTW approach is very appealing because it allows the analyst to estimate the WTW heterogeneity distribution directly (Scarpa et al., 2008) and hence it is the approach we employed in our study. Besides, we used a population of patients waiting for a transplant, which could reduce the chance of poor understanding of the potential choices respondents have to take in the experiment. We find a significant WTW heterogeneity for all the attributes in the experiment and that the WTW correlates with patients' observable characteristics, namely age and time spent on dialysis. Our result implies that patients' welfare may be improved by embedding their preferences into the allocation algorithms.

The remainder of this paper proceeds as follows. Section 2.2 provides details of the experimental procedures and describes the subjects involved in the study; Section 2.3 describes our modelling approach; Section 2.4 presents the results; and finally, Section 2.5 provides the discussion of the results.

2.2 Choice Experiment

Choice experiments (CEs) are stated preference methods useful for eliciting preferences of individuals when revealed preferences cannot be observed. There has been a growing interest in using CEs as an instrument to address a range of health policy questions (de Bekker-Grob et al. 2012; Ryan and Gerard 2003), in particular, to describe ex-ante preferences among different potential treatments. In CEs, individuals are administered with a list of choice sets, including two or more alternatives. The underlying model assumes that the utility of a good is derived not from the good *per se* but from the attributes that the product contains (Lancaster, 1966). Each alternative is therefore defined by values taken by a set of attributes, and individuals are asked to choose the preferred alternative for each choice scenario. The possibility to include continuous variables such as cost and waiting time attributes in the estimations allow researchers to estimate willingness to pay (WTP) (Hole 2008; Nieboer et al. 2010) or WTW (Watson et al. 2004; Brown et al. 2015; Rousseau and Rousseau 2012; Hagemi et al. 2017; Marshall et al. 2018) for variations in attributes' levels. Those measures are meaningful preference parameters if the results of a CE are interpreted within a random utility framework (McFadden 1974; McFadden and Train 2000).

In CEs, the first step involves selecting attributes and levels, the second step is choosing suitable experimental design technique for the choice sets, the third step is recruiting participants and collecting data, and the last step is analysing data through appropriate econometric tools.

2.2.1 Selection of attributes and levels

In this study, attributes and levels were selected in consultation with kidney surgeons at the Kidney and Pancreas Transplantation Unit, University of Padua. For some of the attributes, the levels were determined based on historical data[5]. Two attributes are enumerable (i.e., waiting time and expected graft survival), while the other two (infectious risk and neoplastic risk) being qualitative still, have a precise order in the levels: augmented risk is higher than standard. The attributes and the exact wording used in the experiment were chosen to be familiar to the patients on the waiting list. The kidney surgeons use the same wording and terminology to describe the levels of the risk attributes (standard vs augmented) to patients. Table 2.1 presents a summary of the attributes and levels used in our study.

[5]We had access to the full database of transplants executed in Padua in the last 15 years.

Table 2.1: Attributes and levels used to define the kidney transplant options

Attributes	Definition	Levels
Waiting time	The number of months one has to wait in order to obtain the proposed transplant	6, 12, 36, 60 months
Expected graft survival	The expected length of time the kidney functions well enough to keep recipients from either needing initiation (or return to) dialysis, or another transplant	10, 15, 20 years
Infectious risk	The risk of contracting infectious disease through the transplanted organ	Standard Augmented
Neoplastic risk	The risk of contracting a tumour through the transplanted organ	Standard Augmented

Waiting time is the number of months that patients have to wait to obtain the proposed transplant. This attribute allows patients' to evaluate an approximate waiting time, although there is a chance of waiting lower or higher than declared. The inclusion of the waiting time attribute in our study allowed us to estimate the amount of extra time which a patient is willing to wait for different levels of the other kidney transplant attributes.

The expected graft survival attribute is determined by the features of the organ itself, the characteristics of the recipient and the compatibility between donor and recipient. This attribute allows respondents to understand how long the transplanted organ is functioning. The assessment is the result of a probabilistic calculation based on previous clinical data and experience of the doctor performing the evaluation, but it is subject to a certain degree of uncertainty.

The two risk attributes have two qualitative levels. The wording chosen to describe the levels is the same as those used by the surgeons to describe marginal kidneys to their patients. Standard risk includes cases for which the evaluation process did not identify any risk factor for transmittable disease, and it is the most frequent condition in the assessment of donors and grafts. It is commonly defined as a standard risk to make clear to patients that a zero risk kidney does not exist since infectious or neoplastic pathologies can still be transmitted even if guidelines and good clinical practice are followed. In augmented risk, some of the controls have not been performed, or the donor had some risky behaviours or some kinds of neoplastic disease in the days before his or her death, but an infection or neoplasm may still not result from clinical diagnostics (even if it is possible).

CEs commonly include a cost attribute to be able to compute willingness to pay (WTP) for changes in the composition of a good or service. However, in this study, we did not include a cost attribute as its inclusion is unrealistic within a publicly provided health care system, as in Italy because medical services are typically free at the point of delivery and patients do not have an experience of paying for medical services.

2.2.2 Experimental design

Experimental design is the combination of the attribute levels used to construct the alternatives included in the choice sets. The natural choice would be a full factorial design, namely an experiment that contains all possible combinations of the levels of the attributes. Given the number of attributes' levels in our context, a full factorial design gives rise to 48 possible scenarios ($2^2 * 3 * 4$) that can be combined into 1128 possible choices. Running a CE where respondents are asked to choose from 1128 choices is unfeasible. The standard practice to reduce the dimension of the experimental design is to pick a statistical efficiency measure and select a possible subset of choices that maximise such a criterion. This approach aims to optimise the informativeness of the selected choices, but have no connection with utility theory.

McFadden and Train (2000) showed that to link individual parameters' estimates obtained from a CE to individual preferences, a necessary condition is to assume preferences are complete, monotone and transitive. As a result, we follow Battiston et al. (2016), and we make these assumptions at the design stage. We employed an algorithm that searches for a list of choice sets in which dominant alternatives do not appear, choice sets are not repeated, and the number of choice sets for which the answer can be inferred from the previous one is minimised (assuming transitivity and monotonicity).

We used the 'AlgDesign Package' in R (Wheeler 2006; Aizaki and Nishimura 2008) to generate a D-efficient design of 16 choice sets with D-error of 0.23.[6] We carefully designed the experiment to be consistent with economic theory, allowing us to obtain reliable measures of individual WTW for changes in transplant attributes and to analyse their empirical distribution. In each choice task, patients were asked to select their preferred alternative among two kidney transplantation options, with no "opt out" alternative (see Figure 2.1 for an example choice task).

Although including an opt-out alternative in many stated preference experiments is believed to increase the realism of model estimation, it may also be unnecessary. If individual

[6] We run a pilot taking students as subjects and there was no fatigue effect with 16 choice sets.

preferences are measured to ascertain which components define the most preferred program or treatment, the inclusion of an opt-out alternative might not be a necessity but rather a threat to efficiency. The decision to include an opt-out option is, therefore, determined by the objective of the CE in the first place (Veldwijk et al., 2014). In our study, as patients were enrolled on the waiting list for kidney transplantation, we assumed they would choose from the given transplant alternatives, and hence not receiving a kidney transplant (opting-out) is not considered as a valid choice. Moreover, since we assume completeness, transitivity and monotonicity at the design stage, we did not include an indifference (opt-out) option among the possible answers to the choice sets.

Which of the two treatments would you prefer? (Please put an X below the chosen treatment)

	Treatment A	Treatment B
Waiting Time	36 months	6 months
Graft Survival	15 years	10 years
Infectious Risk	Standard	Augmented
Neoplastic Risk	Augmented	Augmented
Your Choice	☐	☐

Figure 2.1: Translated illustration of a choice task used in the experiment (Original in Italian)

2.2.3 Participant recruitment and ethics

We interviewed 250 patients waiting for a kidney transplant at the Kidney and Pancreas Transplantation Unit, the University of Padua. The interview took place during the periods from April 14, 2015, to June 6, 2017. The transplant centre in Padua is one of the biggest transplant centres in Italy, and it is comparable to other transplant centres outside Italy. We had the unique opportunity of running the experiment on the entire population of individuals waiting for a kidney transplant. The key advantage of administering the questionnaire to this population is that the respondents know precisely the problem and the proposed transplant attributes: as a matter of fact, only two questionnaires were discarded due to item non-response by the patients.

Trained interviewer conducted the face-to-face interviews[7] with a Paper Assisted Personal

[7] The interviewer explained the experiment and obtained informed consent from the participants. We have an added advantage of using the face-to-face interview in that the response rate was 100%.

Interview (PAPI) methodology. In addition to the CE, which is the central part, the questionnaire included some demographic characteristics of recipients used to estimate their effect on the WTW for changes in the time and risk attributes of kidney transplantation. A copy of the survey instrument translated to English is attached in the appendix (original in Italian). Ethical approval for the study was obtained from the Ethical Committee of University Hospital Padua.

Our descriptive analysis indicated that the average time spent on dialysis was about three years. For the major demographic variables, the characteristics of the subjects involved in our study were comparable to the patients in the U.S (Table 2.2). For instance, Matas et al. (2015), based on the Organ Procurement and Transplantation (OPTN) annual report of 2013, indicated that more men than women were on the waiting list (males 60% and females 40%) which is 65 % males and 35% females in our study. The report also indicated that about 35.6% of patients in the U.S. are below 50 years of age, 44% between 50-64 years, and 20.8% above 65 years of age while in our study 45% of them are below 50 years, 45% between 50-64 years and the remaining 10% are above 65 years of age. The same report indicated that about 66% of the patients have a waiting time of fewer than three years, which in our case is about 58%.

Table 2.2: Patients characteristics, external validity

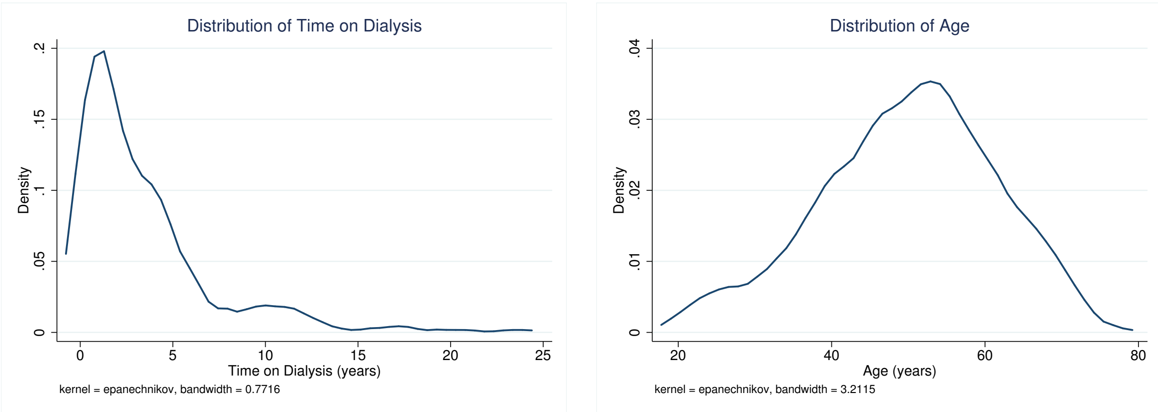
	Padua patients	**US patients**
Gender (males)	0.65	0.60
Age		
below 50	0.44	0.35
50-64	0.45	0.44
65+	0.10	0.21
Blood type		
A	0.30	0.30
B	0.10	0.16
AB	0.03	0.03
O	0.56	0.52

Table 2.3 displays the main characteristics of the candidates. The majority of the candidates were male (65%) with a mean age of 50 years, and 63% of the candidates were employed (working). Also, 48% of them completed high school education, 7% had primary education, and 16% with college/university education. The average time spent on dialysis was about 3.4 years comparable with the national average of 2.8 years (Sefora et al., 2013), and the majority of them (75%) followed haemodialysis while the remaining 25% were on a peritoneal type of dialysis.

Table 2.3: Demographic and clinical characteristics of kidney transplant candidates

	N	%
Characteristic		
Age group		
21-46	84	33.87
46-56	95	38.31
56+	69	27.82
Duration of dialysis		
< 3 years	146	58.87
3-10 years	83	33.47
> 10 years	19	7.66
Health status		
Excellent	6	2.42
Very good	25	10.08
Good	88	35.48
Fair	108	43.55
Poor	21	8.47
Dialysis Modality		
Haemodialysis	181	74.79
Peritoneal	61	25.21
Gender		
Male	162	65.32
Female	86	34.68
Employment status		
Working	155	62.50
Not working	93	37.50
Education		
Elementary	16	6.45
Junior High	74	29.84
High School	118	47.58
College	40	16.13
Number of children		
0	90	36.29
1	57	22.98
2	68	27.42
3	26	10.48
> 3	7	2.82

Figure 2.2 presents the kernel density plots of the distributions of two covariates considered for further analysis. In figure 2.2a, we see a wide variation in the time spent on dialysis, and it is left-skewed with a larger part of the mass between zero and five years. Figure 2.2b shows the variation in age across the patients and the distribution is fairly symmetric perhaps slightly right-skewed with centre around 50 years of age.



(a) Time on dialysis (b) Age

Figure 2.2: Kernel plots of the distribution of covariates (time on dialysis and age)

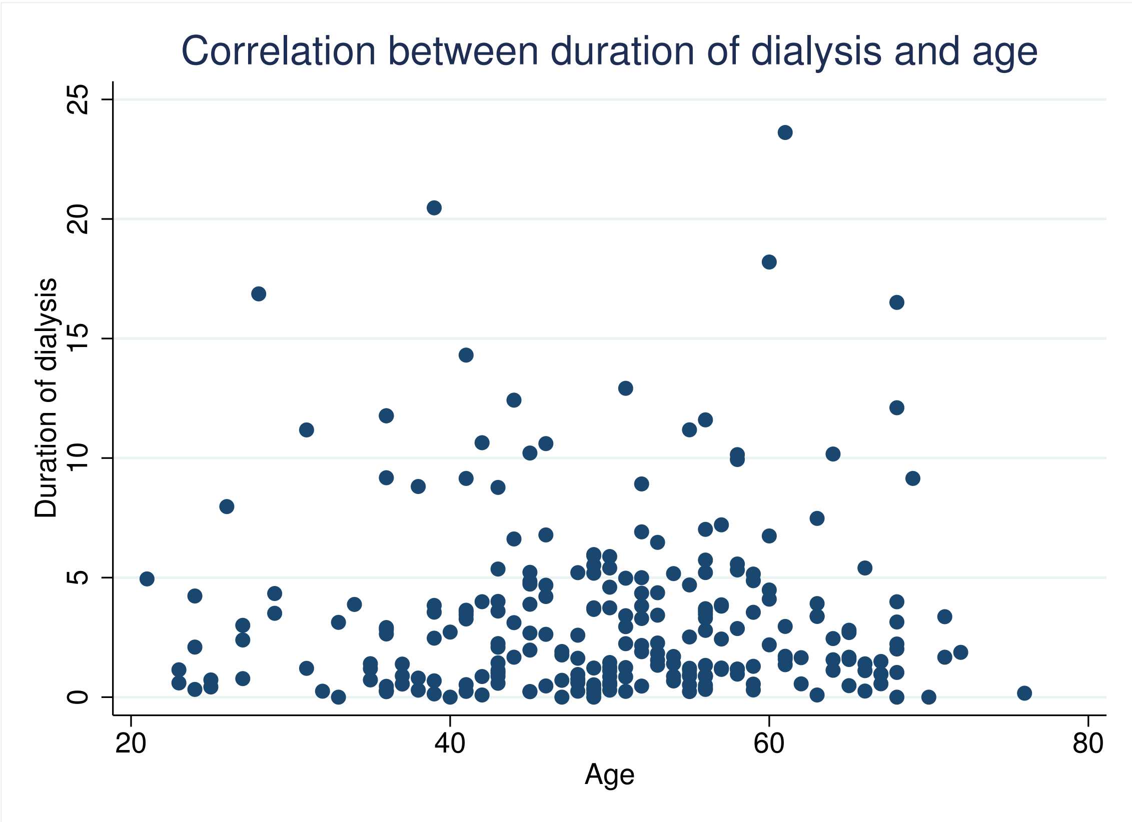


Figure 2.3: Correlation between duration of dialysis and age

Figure 2.3 shows the correlation between age and duration of dialysis. Although patients' age and the length of dialysis may seem strongly associated, the figure shows a very low correlation between the two covariates. It is possible that an older patient might have spent a shorter time on dialysis. In particular, the correlation between age and time on dialysis is around -0.0079. Hence, the effect of age on the WTW for changes in the attributes of kidney transplantation can be interpreted independently of the duration of dialysis.

2.3 Modelling approach

The analysis of responses obtained from a choice experiment is typically modelled using the random utility maximisation (RUM) framework (McFadden, 1974), which is based on three behavioural assumptions: random utility (Thurstone, 1927); multi-attributes utility (Lancaster, 1966); and utility maximisation (Samuelson 1938; Manski 1977). The random utility hypothesis indicates that the utility attached to a good/service has a systematic (or observable) component and a stochastic (or unobservable) component. Lancaster's multi-attribute utility theory indicates that utility is derived not from the product itself but from the attributes that the product contains. The utility maximisation hypothesis stipulates that the patients act rationally and always select the alternative associated with the highest level of utility.

Patients may have heterogeneous preferences for attributes of kidney transplantation over and above the heterogeneity caused by observed individuals' characteristics. Roth et al. (2004) argued that patients have heterogeneous preferences over compatible kidneys. Under mild regularity conditions, any discrete choice model derived from RUM framework has choice probabilities that can be estimated by a mixed logit model (McFadden and Train, 2000). The model, being fully parametric, is adequately flexible that it helps to specify individual, unobserved heterogeneity and allows the researcher to exploit a rich variety of information about behaviour from repeated choices (Greene and Hensher, 2003). Further, the mixed logit model is not subject to the independence of irrelevant alternatives (IIA) assumption, accommodates correlations among panel observations, and accounts for unobserved heterogeneity in tastes across respondents (Özdemir et al., 2009). In mixed logit modelling, we use the estimated preferences (regression coefficients) to compute WTW or WTP values. However, it is possible to directly determine the WTW values (instead of calculating them) and then compute preferences from the estimated WTW values. This procedure is called WTP-space modelling. Since we used a 'waiting time' instead of a 'monetary' attribute, we have a model in WTW space.

2.3.1 Heterogeneity in WTW-space

The mixed logit model estimation requires distributional assumptions on the preference parameters, and the WTW values are computed based on the ratio of the attribute of interest and the waiting time attribute. The distribution of the WTW for a particular attribute then depends on the distributional assumption made on the preference parameters. For instance, if one assumes a normal distribution for the attributes 'expected graft survival' and 'waiting time', the computed WTW for an extra year of expected graft survival, which is the ratio of the coefficient of 'expected graft survival' to the coefficient of 'waiting time', will not have defined moments. To avoid this problem, many studies in this arena usually suggest fixing the denominator, and hence the distribution of the WTW value will take the distribution of the numerator. However, it may be unrealistic to assume that all the patients have the same (fixed) preference for the waiting time. Specifying the continuous attribute (in our case the waiting time attribute) to be log-normally distributed, as suggested by Hole and Kolstad (2012), will help obtain meaningful WTW values with defined moments. However, still, this will result in unrealistic estimates of the means and standard deviation of WTW values and heavily skewed distributions (Hole and Kolstad, 2012). To overcome such a problem, the mixed logit model could be estimated in WTW space rather than preference space (see, for instance, Train and Weeks 2005; Hole and Kolstad 2012).

The utility patient m obtains from choosing kidney transplant alternative t in a choice set s is specified as a function of waiting time, $time_{mts}$, and other attributes of the transplant, X_{mts}:

$$U_{mts} = -\alpha_m time_{mts} + \delta ASC_{mts} + \beta_m' X_{mts} + \varepsilon_{mts} \tag{2.1}$$

where α_m is patient-specific coefficient for waiting time, β_m is a vector of patient-specific coefficients for the other attributes of a kidney transplant, and ε_{mts} is a random term.

The term 'ASC', called alternative-specific constant, captures the systematic effect of the alternative on the left on patients' choices. It is used to measure the systematic tendency of choosing alternative 1 relative to 2 (decision biases). Because the kidney transplant alternatives are unlabelled (i.e. Treatment 1 versus Treatment 2), this effect should not be significant. In other words, we used generic labels for the kidney transplant alternatives. Such generic labels should have no impact on patients' choices as they do not convey additional information about the content of the kidney transplant alternatives. However, in CEs the term 'ASC' is usually found to be significant because of an order effect called left-to-right (reading) bias.

ε_{mts} is assumed to be extreme valued distributed with variance $\mu_m^2(\pi^2/6)$, where μ_m is patient-specific scale parameter. As utility is ordinal, dividing equation 2.1 by the scale parameter allows to obtain scale independent utility (Scarpa et al., 2006). According to Train and Weeks (2005), dividing utility (Equation 2.1) by μ_m does not affect behaviour and results in a new error term, that has the same variance for all decision-makers, which is independently and identically distributed (IID) extreme value distributed with variance equal to $\pi^2/6$:

$$U_{mts} = -(\alpha_m/\mu_m)time_{mts} + (\beta_m/\mu_m)'X_{mts} + \varepsilon_{mts} \tag{2.2}$$

The utility, given the coefficients $\lambda_m = (\alpha_m/\mu_m)$ and $c_m = (\beta_m/\mu_m)$, is written as:

$$U_{mts} = -\lambda_m time_{mts} + c'_m X_{mts} + \varepsilon_{mts} \tag{2.3}$$

The WTW for an attribute is the ratio of the attribute's coefficient to the waiting time coefficient: $WTW = c_m/\lambda_m$. The utility function in the WTW space model (Train and Weeks, 2005) can be written as follows:

$$U_{mts} = -\lambda_m time_{mts} + (\lambda_m WTW_m)'X_{mts} + \varepsilon_{mts} = -\lambda_m(time_{mts} - WTW'_m)X_{mts} + \varepsilon_{mts} \tag{2.4}$$

A theory-consistent reason to estimate the model in WTW space (rather than preference space) is that the variation in WTW must be scale-free, which occurs when the denominator (i.e., waiting time) is no more fixed. If the waiting time coefficient is constrained to be fixed when in fact the scale varies over observations, then the variation in scale will be confounded with the variation in WTW for transplant attributes. Such confounding can be disentangled when we re-parameterise the model such that the parameters are the WTW for changes in each attribute rather than the utility coefficient of each attribute. Therefore, in a setting in which scale can vary over patients, WTW-space model is more appropriate for distinguishing WTW variation from the variation in scale.

Using the WTW-space approach, we assumed a normal distribution on the WTW for an extra year of graft survival and the two risk attributes, but a log-normal distribution on the preferences for waiting time. The approach allowed us to obtain more realistic WTW distributions and overcome the highly skewed distributions that would occur in the preference space model. We estimated two models in the WTW-space. The first is a pooled model (basic mixed logit model in WTW space), where the standard deviation is the same, but the mean differs. The basic model does not take patients' observed characteristics into account. The coefficient of the 'time' attribute is log-normally (LN) distributed with mean $\overline{\lambda}$ and standard deviation σ_λ[8]:

[8]The σ_λ parameters enable quantifying the variability in WTW among patients; a σ_λ parameter significantly different from 0 indicates larger variability in WTW among patients.

$$\lambda_m \sim LN(\overline{\lambda}; \sigma_\lambda)$$

$$\lambda_m = exp(\overline{\lambda} + \sigma_\lambda \eta_m) \tag{2.5}$$
$$\eta_m \sim N(0,1)$$

The WTW for the change in the kth attribute is normally distributed with mean $\overline{WTW_k}$ and standard deviation σ_k:

$$WTW_{mk} \sim N(\overline{WTW_k}; \sigma_k)$$

$$WTW_{mk} = \overline{WTW_k} + \sigma_k \eta_m^k \tag{2.6}$$
$$\eta_m^k \sim N(0,1)$$

2.3.2 Mean heterogeneity in WTW-space model

An important question from a policy perspective is whether the patient-level observable characteristics correlate with patients' WTW for changes in the attributes of kidney transplantation. We propose an approach to represent the role of observable characteristics in explaining the variation in WTW, where such observables can be included as a relevant metric in kidney allocation protocol. Our approach is based on the assumption that the mean of the distribution can be modelled as a function of the socioeconomic variables. We refer to this second model as 'mean heterogeneity in WTW-space' model. Detailed discussion on how to account heterogeneity around the mean of the distribution in the mixed logit framework can be found in Greene et al. (2006) and Bhat (2000). The coefficient of the 'waiting time' attribute is still log-normally distributed with mean $\overline{\lambda}$ and standard deviation σ_λ, but now the mean is a function of the covariates $\overline{\lambda}(age,\ dialysistime)$:

$$\lambda_m \sim LN(\overline{\lambda}(age,\ dialysistime); \sigma_\lambda)$$

$$\lambda_m = exp(\overline{\lambda} + \varphi_1 age + \varphi_2 dialysistime + \sigma_\lambda \eta_m^k) \tag{2.7}$$
$$\eta_m^k \sim N(0,1)$$

The WTW for the kth attribute is normally distributed with mean $\overline{WTW_k}$ and standard deviation σ_k but now the mean is a function of the covariates $\overline{\lambda}(age,\ dialysistime)$:

$$WTW_{mk} \sim N(\overline{WTW_k}(age,\ dialysistime); \sigma_k)$$

$$WTW_{mk} = \overline{WTW_k} + \delta_{1k}age + \delta_{2k}dialysistime + \sigma_k\eta_m^k \tag{2.8}$$
$$\eta_m^k \sim N(0,1)$$

Equation 2.8 is called mean heterogeneity in WTW-space model. We allow WTW_{mk} to vary across individuals both randomly and systematically with observable variables, such as age and time on dialysis. WTW_{mk} is the random willingness to wait for the kth attribute faced by patient m. The term $(\overline{WTW_k}+\delta_{1k}age+\delta_{2k}dialysistime)$ accommodates heterogeneity in the mean of the distribution of the random WTW. Using our data, we estimated models 2.5, 2.6, 2.7, and 2.8.

We employed kernel density plots to show the heterogeneity in WTW for changes in each transplant attribute levels and to examine how WTW varies with observable characteristics. We also presented the cumulative density functions (CDF) of WTW estimates to describe variations in the WTW in term of first-order and second-order stochastic dominance approach.

2.4 Results

2.4.1 Results of the basic mixed logit model in WTW-space

Results are presented in Table 2.4. Patients exhibit a substantial amount of WTW heterogeneity for each of the three attributes of kidney transplantation. The heterogeneity is shown by the statistically significant coefficients of the standard deviation (SD) of the mixed logit model in WTW-space (column 2 of Table 2.4). All the SD parameters are significant, implying the presence of heterogeneity in WTW for changes in the multi-attribute content of kidney transplantation. Each attribute is significant, confirming that all the four attributes are important to patients when evaluating kidney transplant alternatives. As expected, waiting time has a significant and negative effect, meaning that patients, all else being equal, prefer a kidney transplant with a shorter waiting time.

Table 2.4: Basic mixed logit model in WTW-space (normal, waiting time-log-normal)

	(1) (Mean)	(2) (SD)
Waiting time(λ)	-2.7136***	0.9323***
	(0.0890)	(0.1165)
$WTW_{survival}$	5.3148***	4.6941***
	(0.4760)	(0.4116)
$WTW_{standard\ infectious\ risk}$	27.9683***	24.6197***
	(1.9938)	(1.9626)
$WTW_{standard\ neoplastic\ risk}$	27.6704***	21.0165***
	(2.1429)	(2.1205)
ASC	3.4773 ***	-
	(0.6989)	
Number of observations	7936	
Number of respondents	248	
Log-likelihood	-2134.741	
McFadden-R^2	0.193	
BIC	4350.294	

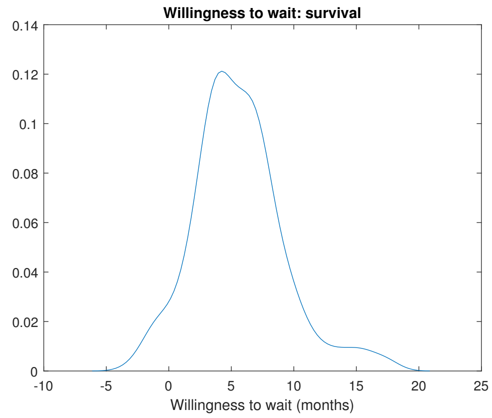
Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

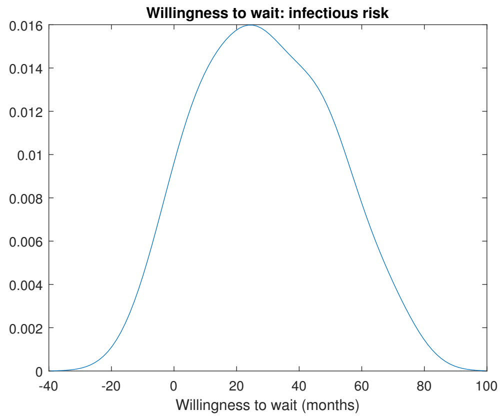
BIC: Bayesian Information Criterion defined as $BIC = -2LL + klog(n)$; ASC: Alternative Specific Constant. The attributes 'waiting time' and 'graft survival' were coded as continuous variables, while 'infectious risk' and 'neoplastic risk' were dummy-coded. In our estimation, the attributes 'expected graft survival', 'infectious risk', and 'neoplastic risk' were normally distributed while the attribute 'waiting time' was log-normally distributed.

The mean WTW for an extra year of graft survival was about five months. In other words, patients are willing to wait, *ceteris paribus*, five months for a kidney which will provide one more year of survival. On average, patients are willing to wait, *ceteris paribus*, 27 months more for a kidney of standard risk (both infectious and neoplastic) as compared to one of augmented risk. These are mean WTW estimates, which show the average WTW for a particular level of the transplant attribute in the population of interest. The statistically significant coefficient of ASC indicates the presence of left-to-right biases in our data. Except for accounting left-right (decision) biases, the ASC has no additional meaning as the kidney transplant alternatives were unlabelled or generic.

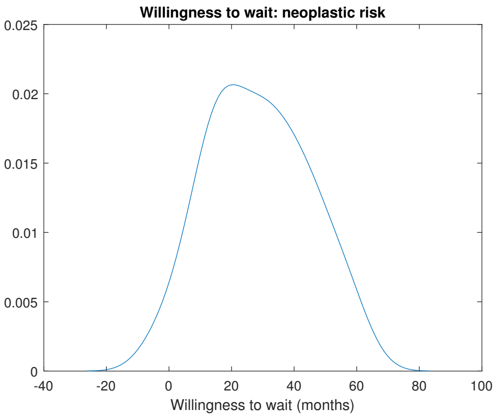
We explored the entire distribution of WTW using kernel density plot (Figure 2.4). These plots show the presence of WTW heterogeneity among patients. The distribution of WTW for an extra year of graft survival presented in panel 2.4a indicates heterogeneity in WTW: the distribution is concentrated at about five months, but on the right side it has a long tail implying the presence of a fraction of patients who are willing to wait much longer than the average WTW. 125 (50%) of the patients are willing to wait above five months for a transplant that will provide an extra year of survival. In figure 2.4b and 2.4c, the distributions are less concentrated compared to figure 2.4a, which indicated more heterogeneity in WTW for changes in the risk attributes. Similarly, the distribution of individual WTW for neoplastic risk presented in 2.4c, shows variations in WTW, which indicates the presence of heterogeneity in the valuation of neoplastic risk. Although the attribute 'waiting time' is log-normally distributed, we still see parts of the WTW distribution that show people preferring higher risk and lower expected survival. These results might have occurred because the other attributes were specified to be normally distributed resulting in the negative parts of the WTW distributions. In sum, both the basic mixed logit in WTW-space model and the kernel density plots showed the presence of heterogeneity in patients' WTW for changes in the transplantation attributes.



(a) WTW for extra year of survival



(b) WTW for standard infectious risk



(c) WTW for standard neoplastic risk

Figure 2.4: Kernel density plots of the distribution of individual WTW

We also performed further analysis to aid further interpretation and generalisability of the results. We estimated the pooled/basic mixed logit in WTW-space model to recalibrate for 5-year graft survival differences. The results are presented in Table 2.5 and Figure 2.5 show the distributions of WTW. The variable '$WTW_{survivalof15years}$' relates to the average

WTW for a kidney that offers 15 years of survival rather than 10 years. The benchmark for comparison is an organ which will provide 10 years of survival. On average, patients are willing to wait, *ceteris paribus*, 13.7 months more for a kidney that will offer 15 years survival rather than 10 years. The WTW increases for a kidney that will provide 20 years of graft survival compared to 10 years. On average, patients are willing to wait 40 months, all else being equal, for a kidney that will offer 20 years of survival compared to 10 years. The WTW increases by 25 months for changes in the expected graft survival from 15 to 20 years, which is consistent with the WTW of 5 months for an extra year of graft survival.

In the model recalibrated for a 5-year difference in the expected graft survival (Table 2.5), we find that patients are willing to wait some 28 months longer for a kidney with standard infectious risk than for an organ with augmented infectious risk, keeping all other factors constant. Besides, patients are willing to wait some 24.8 months longer for a kidney characterised by standard neoplastic risk than an organ with an augmented risk.

Table 2.5: Basic mixed logit model in WTW-space (normal, waiting time-log-normal)

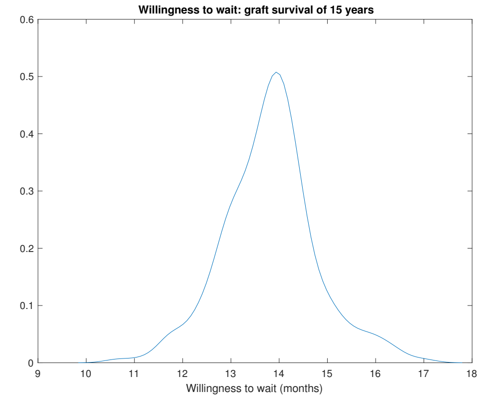
	(1) (Mean)	(2) (SD)
Waiting time(λ)	-2.7591***	1.3975***
	(0.1183)	(0.1701)
$WTW_{survival of\ 15\ years}$	13.7827***	4.2071***
	(1.4839)	(1.3732)
$WTW_{survival of\ 20\ years}$	38.8613***	14.1993***
	(2.2385)	(1.8244)
$WTW_{standard\ infectious\ risk}$	28.2017***	21.6893***
	(0.9707)	(1.6546)
$WTW_{standard\ neoplastic\ risk}$	24.7673***	-18.0375***
	(1.4063)	(1.6857)
ASC	1.6857 ***	-
	(0.519)	-
Number of observations	7936	
Number of respondents	248	
Log-likelihood	-2139.54	
McFadden-R^2	0.269	
BIC	4377.851	

Standard errors in parentheses

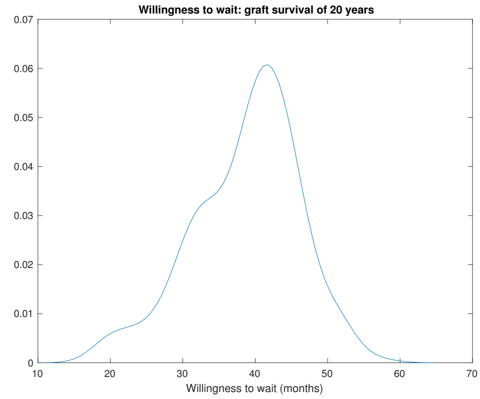
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

BIC: Bayesian Information Criterion defined as $BIC = -2LL + klog(n)$; ASC: Alternative Specific Constant. The attributes 'waiting time' was coded as a continuous variable, while 'infectious risk', 'neoplastic risk', and 'graft survival' were dummy-coded. In our estimation, the attributes 'expected graft survival', 'infectious risk', and 'neoplastic risk' were normally distributed while the attribute 'waiting time' was log-normally distributed.

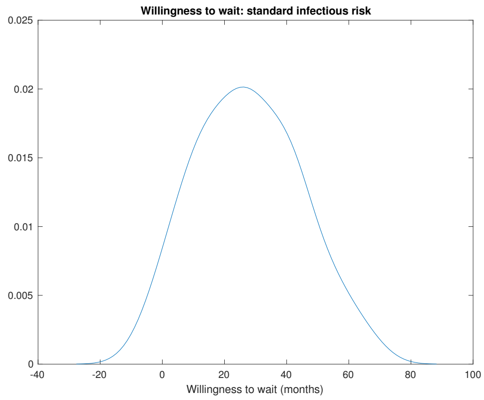
The distribution of WTW for 15 years of expected graft survival presented in panel 2.5a indicates heterogeneity in WTW: the distribution is concentrated at about 14 months. In figure 2.5b, the distributions are more dispersed compared to figure 2.5a indicating more heterogeneity in WTW for 20 years of graft survival than for 15 years.



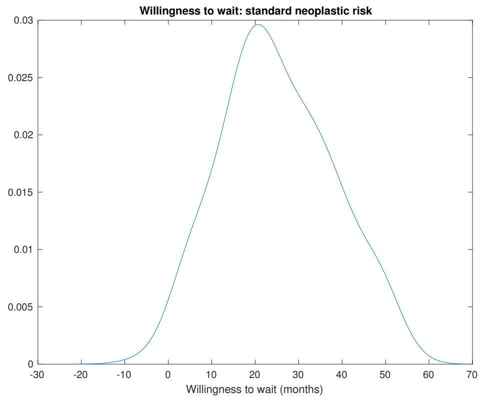
(a) WTW for 15 years of graft survival



(b) WTW for 20 years of survival



(c) WTW for standard infectious risk



(d) WTW for standard neoplastic risk

Figure 2.5: Kernel density plots of the distribution of individual WTW

We extended our analysis to show how accounting for patients' observable characteristics such as age and length of time on dialysis change the WTW estimates. We also included other observable characteristics of the patients (such as education, gender, the presence of children, employment status and so on) to explain heterogeneity in WTW. However, the WTW estimates do not differ across patients based on these covariates. We observe statistically significant differences in WTW across patients only for age and duration of dialysis. Sections 2.4.2 and 2.4.3 provides results of the extended mixed logit model (the mean heterogeneity) in WTW-space.

2.4.2 The effects of age

Due to life expectancy differences, the WTW for changes in attributes of a kidney transplant may vary according to the patients' age. Column 3 of table 2.6 shows the effect of age on WTW for the attributes considered, and all the coefficients of the age effect (δ_{1k}) are statistically significant at the conventional level. The results indicate that the value patients place on a specific attribute of kidney transplant varies by age. All else equal, an extra year of age negatively affected the WTW for change in attributes of a kidney transplant. In particular, a patient with an additional year of age is willing to wait 0.1 months (3 days) less for a kidney that offers an extra year of graft survival. Similarly, all else being equal, a patient with an extra year of age is willing to wait 0.3 months (9 days) less for standard infectious risk and about 0.4 months (12 days) less for standard neoplastic risk.

Table 2.6: Extended mixed logit model in WTW-space (normal, waiting time-log-normal)

	(1) (Mean)	(2) (SD)	(3) (Age effect) (δ_{1k})	(4) (Time on dialysis effect) (δ_{2k})
Waiting time (λ)	-3.4325***	0.9818 ***	0.0170 **	-0.0463 *
	(0.4267)	(0.1240)	(0.0081)	(0.0249)
$WTW_{survival}$	9.1975***	4.2731***	-0.1012***	0.4278***
	(1.7709)	(0.3260)	(0.0330)	(0.0871)
$WTW_{standard\ infectious}$	36.5998***	24.2340***	-0.2603**	1.5848***
	(7.2230)	(1.8752)	(0.1300)	(0.2807)
$WTW_{standard\ neoplastic}$	42.5840***	21.2400***	-0.3594**	1.0195 ***
	(9.2718)	(1.9342)	(0.1644)	(0.3818)
ASC	3.3633 ***			
	(0.6522)			
Number of observations	7936			
Number of respondents	248			
Log-likelihood	-2124.516			
McFadden-R^2	0.197			
BIC	4401.67			

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The coefficient of the 'waiting time' attribute was specified to be log-normally distributed since every patient was expected to prefer a shorter waiting time. The coefficients of the other attributes were assumed to follow a normal distribution. We call the model, an 'extended mixed logit model in WTW space' because it accounts for patients' observed characteristics. BIC: Bayesian Information Criterion defined as $BIC = -2LL + klog(n)$.

The mean heterogeneity in WTW-space model indicates that patients, *ceteris paribus*, are willing to wait 9 months for a kidney that offers an additional year of graft survival.

A patient with an age of 21 years (the minimum in our data) and an average dialysis duration of 3.4 years is willing to wait 8.5 months for a kidney that will last one more year. An older patient (say 65 years old) who spent an average dialysis time of 3.4 years is willing to wait 4 months for a kidney that provides one more year of survival. Other factors being equal, there is a 5-month difference in WTW for an additional year of graft survival between patients who are 21 and 60 years old although the difference in WTW for consecutive ages seem negligible. Table 2.7 shows the WTW for the changes in transplantation attributes (i.e., for a kidney that offers an additional year of survival and an organ of standard risk).

Table 2.7: WTW for different age levels and for an average duration of dialysis (3.4 years)

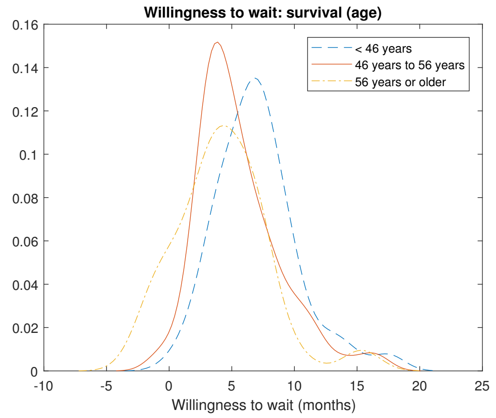
	(1) $(WTW_{extrasurvival})$	(2) $(WTW_{standardinfectious})$	(3) $(WTW_{standardneoplastic})$
Age (years)			
21	8.519	36.494	38.485
25	8.114	35.453	37.047
30	7.608	34.151	35.250
35	7.102	32.850	33.453
40	6.596	31.548	31.656
45	6.090	30.247	29.859
50	5.584	28.945	28.062
55	5.078	27.644	26.265
60	4.572	26.342	24.468
65	4.066	25.041	22.671
70	3.560	23.739	20.874
75	3.054	22.438	19.077
80	2.548	21.136	17.280

Our findings show that younger patients are willing to wait longer for a kidney transplant characterised by better levels of the attributes (i.e., an extra year of graft survival, standard infectious risk, and standard neoplastic risk) compared to older patients.

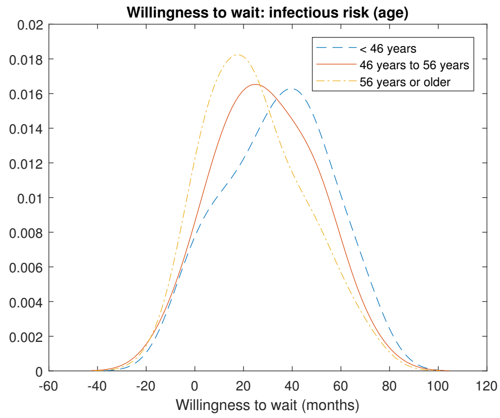
The distributions[9] of the WTW for each of the three attributes are shown in Fig 2.6. In fig 2.6a, the distributions of WTW for changes in each attribute across the three age groups differ, and for patients with age 56 years and above the entire distribution is shifted to the left. For an extra year of graft survival, the distribution of WTW is more dispersed among older patients (56+ years) than the other groups of patients. In figures 2.6b and 2.6c, the

[9]We used kernel density plots to show the distributions of WTW for changes in each of the three attributes. The plots are generated for three age groups: less than 46, 46-56, and +56 years of age.

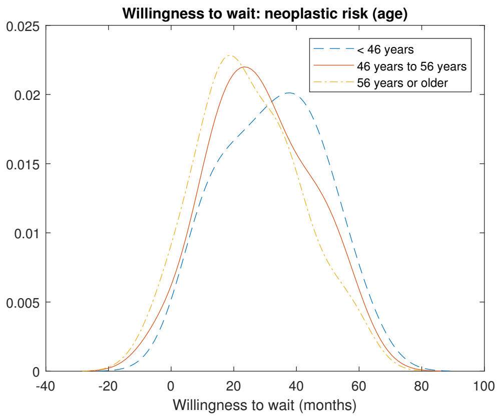
whole distribution of WTW for a transplant with standard risk attributes among patients of 56 years and above is shifted to the left and becomes more concentrated implying less variability in WTW among older patients.



(a) WTW for extra year of survival



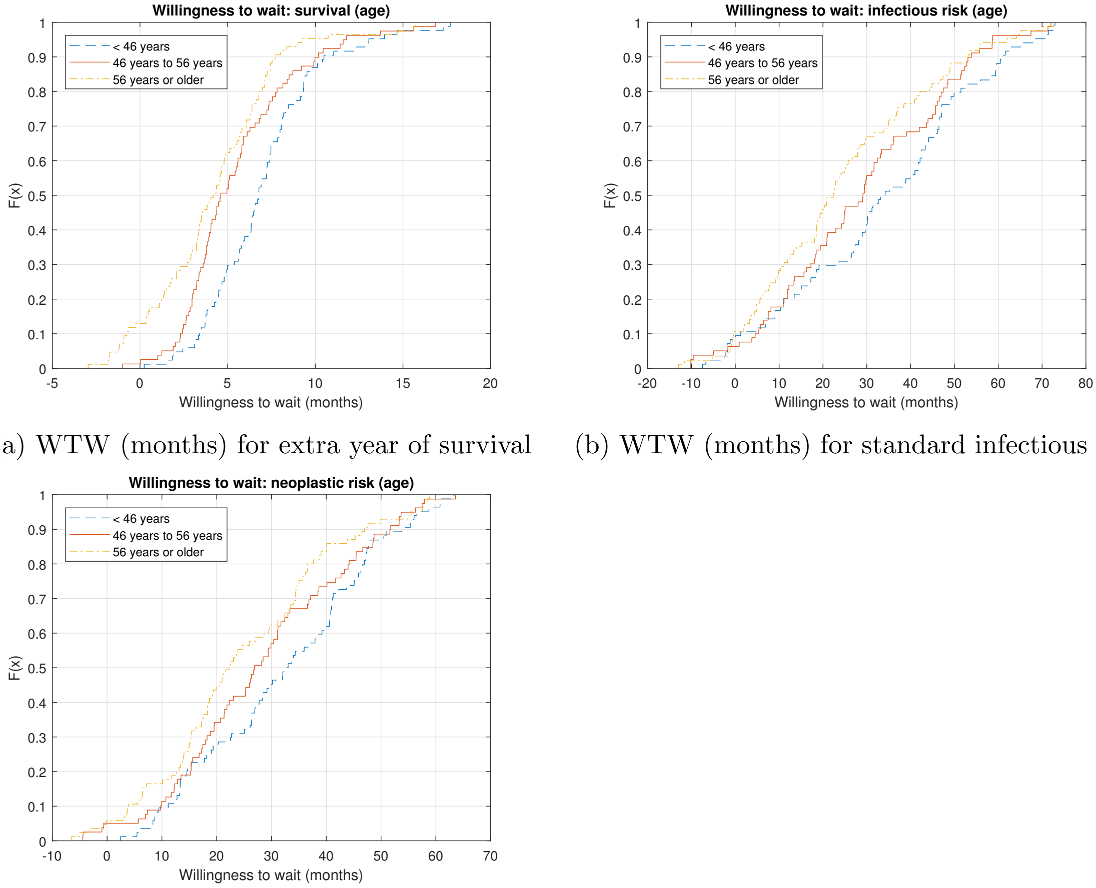
(b) WTW for standard infectious risk



(c) WTW for standard neoplastic risk

Figure 2.6: Kernel density plots of the distribution of WTW: effect of age

In figure 2.7, we show the cumulative density functions (CDFs) of the WTW for changes in each of the three attributes. The plots demonstrate that the WTW for each attribute among patients in the first two age groups (<46 and 46-56 years of age) first-order stochastically dominates the older groups (+56 years). The first-order stochastic dominance provides evidence that for a given initial level of WTW, the probability that WTW exceeds the initial WTW, is higher among the younger patients than, the older ones. For example, given an average WTW for standard infectious risk, the probability that WTW outweighs the average is higher among the younger patients than, the older ones. This evidence suggests that an increase in age is expected to shift the distribution of WTW to the left, and hence a lower WTW. The may imply that keeping a patient on the waiting list as he/she ages may change preferences and thus the WTW. However, accounting for the dynamics in preferences and WTW as age increases need observing a patient at two points in time.



(a) WTW (months) for extra year of survival (b) WTW (months) for standard infectious risk

(c) WTW (months) for standard neoplastic risk

Figure 2.7: Visual representations of the CDF of WTW values: effect of age

2.4.3 The effects of dialysis duration

Patients with irreversible chronic kidney failure and without access to preemptive transplantation need to undergo dialysis treatment while they are waiting for kidney transplantation. The length of stay on dialysis depends, among other factors, on initial health condition, the quality of health care, availability of compatible kidneys, and how much the patients take an active role in their healthcare. The WTW for the change in the levels of each attribute may differ according to the time spent on dialysis[10].

Results are presented in column 4 of Table 2.6. The coefficients of 'time on dialysis effect' on all the attributes are positive and significant, indicating that patients, all else being

[10]The length of time on dialysis was obtained by taking the difference between the date of interview and the starting date of dialysis.

equal, are willing to wait longer for transplantation with a better-expected outcome, the longer the time spent on dialysis. More specifically, a patient who spent an extra year on dialysis, all else being equal, is willing to wait an additional 0.4 months (12 days) for an extra year of graft survival. Similarly, all else equal, a patient who spent an extra year on dialysis is willing to wait 1.6 more months for a kidney transplant with standard infectious risk, and a month for standard neoplastic risk. The mean heterogeneity model in WTW-space highlighted that for each of the three attributes, patients with longer time on dialysis are willing to wait longer for transplantation with a better-expected outcome (i.e., a kidney transplant with an extra year of graft survival, standard infectious risk, and standard neoplastic risk).

The model indicates that for a patient with an average age of 50 years and dialysis duration of 0.5 years, the WTW is 4.4 months for a kidney that will offer one more year of survival. A 50 years old patient who spent a dialysis time of 5 years is willing to wait 6.3 months for a kidney that provides one more year of survival. Table 2.8 shows the WTW for the changes in transplantation attributes (i.e., for a kidney that offers an additional year of survival and an organ of standard risk) for different values of dialysis duration.

Table 2.8: WTW for different duration of dialysis and for an average age (50 years)

	(1) $(WTW_{extrasurvival})$	(2) $(WTW_{standardinfectious})$	(3) $(WTW_{standardneoplastic})$
Duration of dialysis (years)			
0.5	4.351	24.377	25.124
1	4.565	25.170	25.634
2	4.993	26.754	26.653
3	5.421	28.339	27.673
4	5.849	29.924	28.692
5	6.277	31.509	29.712
6	6.704	33.094	30.731
7	7.132	34.678	31.751
8	7.560	36.263	32.770
9	7.988	37.848	33.790
10	8.416	39.433	34.809
11	8.843	41.018	35.829
12	9.271	42.602	36.848
15	10.555	47.357	39.907
20	12.694	55.281	45.004

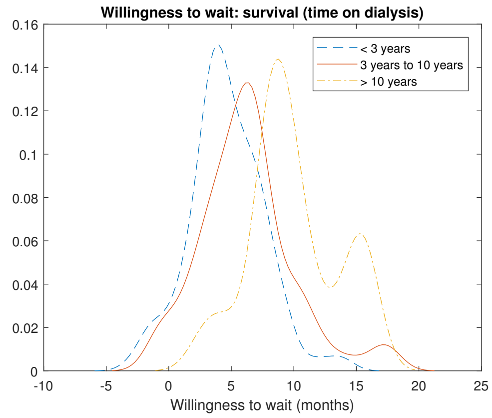
We also explored the distribution of WTW across three groups of patients based on the duration of dialysis[11]. The shapes of the distributions of the WTW are different across

[11]In order to graphically explore the possible variations in WTW according to the variation in the

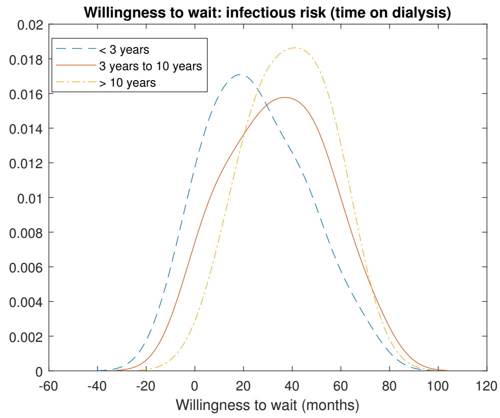
patients with a different length of time on dialysis (Figure 2.8). The distributions of WTW for changes in each of the attributes are shifted to the left among patients who spent over ten years on dialysis. For this group of patients, there is a lower frequency at the mean but a wider distribution elsewhere implying more heterogenous WTW values. While the dispersions are more or less the same for standard infectious risk and standard neoplastic risk, the distribution for an extra year of graft survival is more concentrated. However, for patients who spent less than three years on dialysis, the distributions are moved to the left for all the attributes, which may explain the presence of impatience (time-discounting) mainly at the early stage of dialysis. The distributions suggest the presence of WTW heterogeneity that can be explained by the variation in the time on dialysis.

---

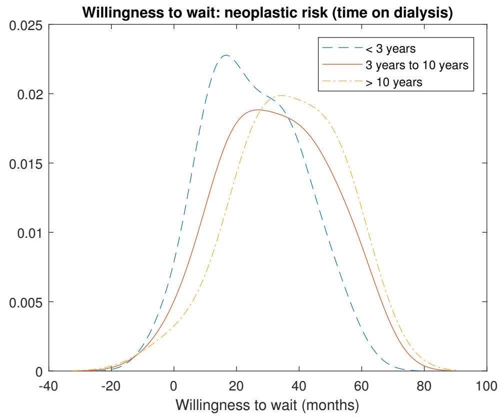
time on dialysis, we considered three groups of patients: those who spent less than 3 years, 3 to 10 years, and over 10 years on dialysis. The data revealed that 58.87% (146 patients) spent less than 3 years, 33.47% (83 patients) spent 3-10 years, and the remaining 7.66% (19 patients) spent above 10 years on dialysis. Using mixed logit in the preference-space model, we performed a subgroup analysis to aid further interpretation and generalisability of the results. We defined the subgroups a priori according to patients' age and duration of dialysis. We classified the different subgroups using 'xtile' command in STATA in which every subgroup contained an equal number of patients. The results are consistent with the classification of patients into subgroups based on age and duration of dialysis shown in the paper. The results are available upon request.



(a) WTW for extra year of survival



(b) WTW for standard infectious risk

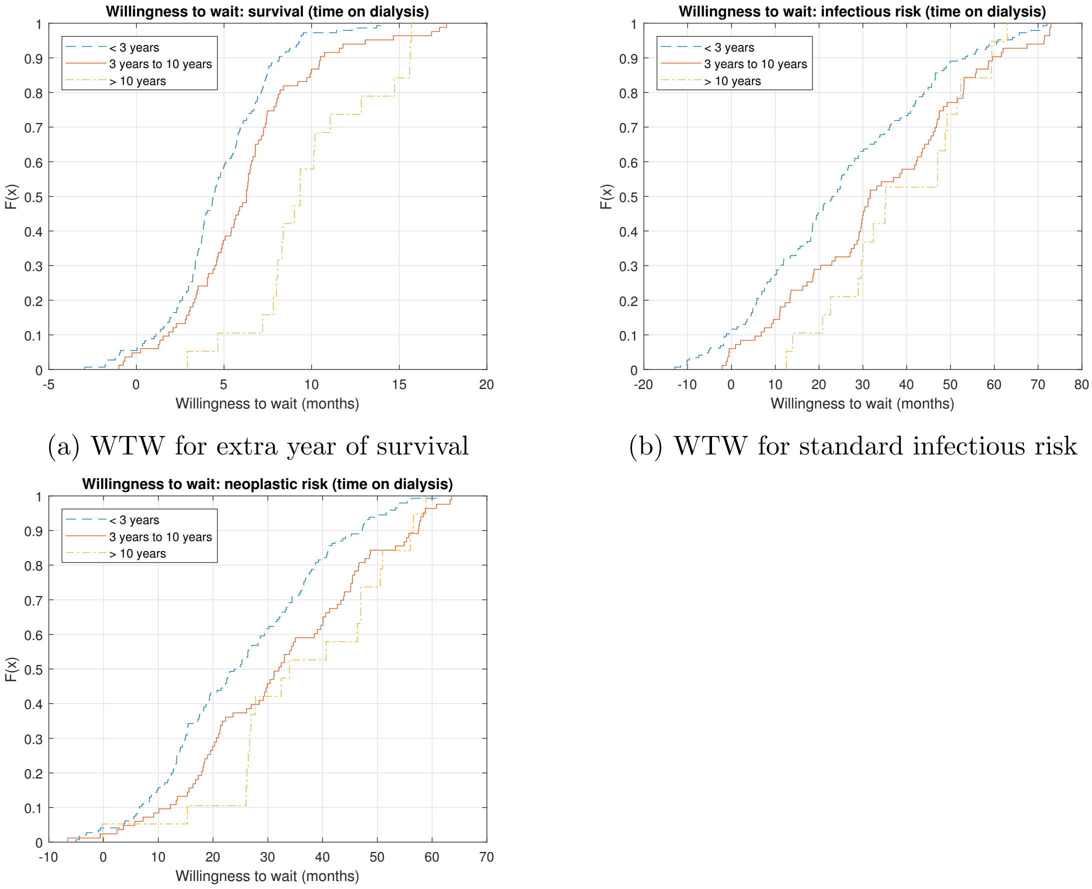


(c) WTW for standard neoplastic risk

Figure 2.8: Kernel density plots of the distribution of WTW: effect of dialysis duration

Figure 2.9 shows the CDF of WTW for the change in each of the three attributes. The CDF of WTW for changes in each attribute among patients who spent over three years of dialysis first-order stochastically dominates those patients with less than three years. At any initial level of WTW, the probability that WTW exceeds that initial level of WTW is higher among patients who spent over three years on dialysis. Panel 2.9a of figure 2.9 implies that for given WTW for an extra year of graft survival (say, five months), the probability that the WTW exceeds five months is higher among patients who spent 3-10 and over ten years on dialysis than those who spent only less than three years. The result raises a question: *shall the decision makers change the way kidney is allocated as patients' time on dialysis increases?* The answer to this question is far from evident unless we observe each patient at two points in time. However, it may suggest that regardless of the initial level of WTW, keeping patients longer on dialysis may increase their WTW. It also means that under weak regularity conditions on the utility function, regardless of the initial heterogeneous level of WTW, keeping patients longer on dialysis may change their preferences and hence a point-by-point shift of the CDF in favour of higher WTW for better kidney transplantation.

Restricting our analysis only to those patients within 3-10 and beyond ten years on dialysis, the first-order stochastic dominance[12] fails for the risk attributes since there is a point at which the two distributions intersect in each plot. However, assuming risk aversion, we observe second-order dominance. The plots suggested that for patients over ten years on dialysis, the distribution of WTW for standard risk attributes of a kidney second-order stochastically dominates those who spent 3-10 years.



(a) WTW for extra year of survival

(b) WTW for standard infectious risk

(c) WTW for standard neoplastic risk

Figure 2.9: Visual representations of the CDF of WTW values: effect of dialysis duration

2.5 Discussion and conclusions

In this paper, we applied a choice experiment to determine the willingness to wait (WTW) in the context of kidney transplantation. We explored patients' WTW for changes in the

[12]First-order stochastic dominance implies second-order stochastic dominance but not necessarily the reverse. In second-order dominance, the CDFs can cross.

duration and risk attributes of kidney transplantation. The result of the mixed logit model in WTW-space supported the existence of WTW heterogeneity for changes in the multi-attribute content of kidney transplantation. By incorporating observable characteristics of the patients in the mean heterogeneity WTW-space framework, we determined the factors that correlate with the WTW for transplant attributes. In the mean heterogeneity WTW-space model, we do observe an impact on the WTW for a kidney transplant attributes with respect of age and duration of dialysis. Considering the impact of age, we find that younger patients are willing to wait longer compared to older patients for a kidney transplant that will offer an additional year of graft survival, and for a kidney characterised by standard and infectious risk. However, when it comes to duration of dialysis, the patients with longer time on dialysis are willing to wait longer for a kidney transplant with a better-expected outcome.

We show that younger patients are willing to wait longer than older patients for a kidney transplant with a better-expected outcome. In this respect, the ECD or the 'marginal' kidneys which otherwise would be discarded may be allocated to older patients while offering the standard criteria donor (SCD) organs to younger patients. This is because the shorter graft survival of the marginal kidneys may be less problematic for older patients who have a relatively shorter life expectancy. Importantly, this may imply that given a shorter life expectancy, older patients could benefit from the marginal kidneys leading to higher utilisation of marginal organs, an increasing number of older patients will get transplants and hence reducing the waiting time for older patients.

Our findings can be considered alongside some earlier non-DCE studies. The implications of our results, based on stated preference experiment, are consistent with previous studies that used administrative data. For instance, using data from the Scientific Registry of Transplant Recipient database and based on survival models, Schold and Meier-Kriesche (2006) have shown that older patients (those 65 years and above) had longer life expectancy when they accepted an ECD within 2 years of ESRD onset (5.6 years) compared with waiting for a standard kidney (5.3 years) or a living donation (5.5 years) after 4 years of dialysis. The authors also indicated that younger patients (18-39 years old) had longer life expectancy with a living donation (27.6 years) or standard kidney (26.4 years) after four years on dialysis compared with an ECD after two years of dialysis (17.6 years).

Many kidney allocation protocols prioritise patients based on the waiting time criteria (first come, first transplanted). Patients who stayed longer on the waiting list end up being prioritised, but less-than-perfect kidneys are typically discarded. This raises the issue of equity-efficiency trade-off. From an equity perspective, patients with longer duration of dialysis may be prioritised, even if someone else who has not spent as long obtains considerable health benefits from transplantation. However, from an efficiency point of view, kidneys could be transplanted to patients deriving greater health benefits, although they are at the early stage of dialysis treatment. Currently, in Padua Transplant Unit, patients with longer duration of dialysis are given the priority for the available compatible kidneys, which is a common feature of most transplant protocols. However, our analyses demonstrate that patients with longer time on dialysis are willing to wait longer for a transplant with better outcomes than patients at the early stage of dialysis. This can be explained by the fact that dialysis may initially involve difficulty and hardship in many respects, but with everyday use and as time passes, dialysis may be more of day-to-day activity. Thus, for patients with longer duration of dialysis, the decision to wait longer for better quality may reflect a system of getting used to the dialysis treatment. However, impulsivity may be an essential factor at the earlier stage of dialysis through the inability to adapt to the dialysis procedure and hence patients at the early stage of dialysis may exhibit a higher degree of time-discounting compared to patients who spent longer time on dialysis. In other words, the inconvenience of dialysis, especially at the beginning, might have created a higher demand for kidney transplantation, and this may have caused a higher WTW-discounting. However, further research is required to understand the mechanism behind this result.

The stochastic dominance approach indicated that at an initial level of WTW, the probability that WTW for better quality kidney exceeds some initial level of WTW is higher among patients with longer time on dialysis, and among younger patients. As time goes by, a question of changing the kidney allocation mechanism may be expected. That is, *should the decision makers improve the way kidney is allocated as patients' time on dialysis increases or as patients' age increases?* The answer to the question is not so apparent as it requires observing each patient at different points in time. However, it may suggest that regardless of the initial level of WTW, keeping patients longer on dialysis or keeping patients on the waiting list as they age may change their preferences and WTW. Further research is required to understand the dynamics of patients' preferences and WTW as age or duration of dialysis increases.

Our results have implications for medical practice and decision makers. The presence of WTW heterogeneity could potentially be relevant at the design stage of future kidney allocation algorithms. The mean heterogeneity in WTW-space model identified age and duration of dialysis as the key observable dimensions that correlate with WTW and thus could be included in the protocol to improve the allocation mechanism. Allocating 'marginal kidneys' to older patients may improve the matching process and could be beneficial if transplanted at the early stage of ESRD. In this regard, an allocation system that aims to increase the number of kidneys from older donors (marginal kidneys) allocated to older recipients may maximise patients' welfare if such systems result in lower waiting time for older recipients. Given a higher WTW-discounting, especially at the early stage of dialysis, offering available compatible kidneys (including 'marginal' organs) even to those patients who are in the early stage of dialysis may improve the matching process, reduce organ wastage, and overall welfare of the patients on the waiting list. Offering the high-quality kidneys to patients with longer duration of dialysis, as long as they are on the waiting list, may bring greater health benefits.

These results must be considered in light of some limitations. The study only looked at the preferences of patients, and kidney surgeons' preferences were not incorporated. Only patients' preferences may not always be used in guiding medical policies. The surgeons' preferences may be against patients' preferences. A choice through shared-decision making, therefore, requires the involvement of both surgeons and patients. Future research should include the kidney surgeons to take part in the decision.

Patients were asked to attend two risk attributes and two duration attributes in the choice of treatments in kidney transplantation. There are likely to be many implicit and often subconscious rules being adopted to process the attributes and alternatives that are used, possibly to cope with the amount of information to assess. Moreover, there may be difficulty in understanding the differences between the two risk attributes as they are nearly indistinguishable in terms of the levels: standard and augmented. Instead of differentiating between the two risk attributes, patients may have used some form of information processing strategies, possibly due to the similar levels assigned to the risk attributes. Further research is required to understand how patients process and understand the risk attributes. Moreover, future research should include supplementary or debriefing questions to better understand patients' knowledge of the choice tasks and the risk attributes.

Discarding kidneys because of their being offered to patients who decline them because of their riskiness of a tumour or infection or the expected graft survival is a major issue. Since an opt-out option was not included in the experiment, the predicted probability of accepting of patients of certain characteristics for a kidney of certain characteristics cannot be computed. Using a forced choice format which, while being relevant to policy, does not allow patients to indicate that they did not prefer one transplant alternative over the other. The inclusion of an opt-out option, however, would have affected the number of choices to the extent that response rate might have been negatively affected. Future study may include an opt-out option if the prediction of uptake or probability of accepting a kidney of certain characteristics (such as the marginal kidney) is required. More research is needed to obtain the exact welfare measures, which require including a constant comparator or status-quo option.

In summary, accounting for WTW heterogeneity could potentially improve the kidney allocation mechanism, patient satisfaction, and general welfare. Older patients would benefit from the marginal organs in the early stage of ESRD onset, while younger patients would benefit from SCD organs even with longer dialysis exposure. The equity-efficiency trade-off may be compromised in some sense if the allocation mechanism could bring greater health advantages to patients on the waiting list.

Chapter 3

Does cognitive ability affect choice consistency?

ABSTRACT

This paper investigates whether there is a link between cognitive ability, choice consistency, and willingness to wait (WTW), using a population of roughly 250 patients enrolled on the waiting list for kidney transplantation in Italy. Patients participate in a choice experiment (CE) measuring time and risk preferences in kidney transplantation and answer a three-item cognitive ability test. Heteroskedastic and generalised multinomial logit models were employed to investigate the effect of cognitive ability on choice consistency. A higher cognitive ability tended to result in more consistent choices and a smaller weight of the error term in the estimated linear-additive underlying utility function. Patients with a higher cognitive ability can state decisions that are based more firmly on the content of the choice tasks, namely the attributes and their respective levels. Further, they tend to provide reliable choice responses and have a lower WTW (more impatient) for a kidney transplant with better-expected outcomes. These relationships are significant, and even when one takes account of other differences between these patients in terms of their age, education, and gender. The study highlighted the importance of incorporating a cognitive ability test in CEs to determine the consistency and hence the reliability of choice responses. Further research in different settings is required to confirm these results. [1]

[1]This is single authored paper. I wish to thank Giacomo Pasini and Luca Corazzini for helpful comments. Financial support of the 'Progetto di Ateneo KIDNEY' from the University of Padua is gratefully acknowledged.

3.1 Introduction

Often, many decisions in life require choosing between alternatives that vary along several dimensions, such as the choices of saving versus consumption, private versus public health insurance schemes, commuting by bus versus train and so on. Variation in these factors impacts the choice decisions of different people differently; for instance, some individuals are more risk-loving, whereas others are more prudent in their choices of consumption versus saving. When making decisions involving choice, people must generally take into account numerical information (consider stock prices, earthquake risks, transplant risks, calorie counts), but not all individuals have the capacity to understand and use numbers. Instead, individuals differ concerning their cognitive abilities and skills, and such differences can predict logically consistent decisions. Understanding cognitive abilities are essential for economists at least for two reasons. First, cognitive ability is crucial for decision-making as it influences individuals' ability to process information and to make the right choices. Second, cognitive functioning may be considered as one aspect of human capital, along with health and education (Mazzonna and Peracchi, 2012).

Studies in psychology and economics have recently tried to understand better if and how cognitive abilities affect behaviour across a wide range of tasks. Notable examples include studies on economic choices (Benjamin et al. 2013; Deck and Jahedi 2015; Dohmen et al. 2010; Frederick 2005) and and strategic interaction games (Cappelletti et al., 2011). Although cognitive ability and economic behaviour have become a topic of considerable research interest and policy concern, the effect of cognitive ability on response consistency and the level of cognitive ability needed to consistently complete choice questions in stated preference experiments is currently under-researched. A possible problem with this so-called stated preference approach is that individuals may provide random responses in choice decision-making, especially when they are confronted with several choice tasks. An important question from a policy perspective is whether people are consistent in decisions that involve choices as in choice experiments (CEs).

CEs, a stated preference elicitation method, are frequently used in applied economics to measure individuals' preferences for various aspects of non-market goods, e.g., in health care, how do patients value and trade-off factors such as treatment effectiveness and risk of side effects in the delivery of healthcare (de Bekker-Grob et al. 2012; Ryan et al. 2008; Ryan et al. 2007). In CEs, respondents are presented with choices sets, each of which contains two or more alternatives that vary with respect to attribute levels. For each choice task, respondents are expected to face trade-offs between attributes and based on these trade-offs, they state what alternative they would choose. However, the consistency of the choice responses is likely to depend on the proper evaluation of each piece of information and the ability of respondents in making non-random choices. In this regard, greater cognitive ability is often considered an essential requirement for non-random and consistent decisions.

In this study, I test the hypothesis that response consistency in decisions involving choices is related to cognitive abilities. More specifically, this study aims to investigate whether patients' cognitive ability potentially affects the choice consistency and willingness to wait (WTW) in a CE concerned with time and risk preferences for kidney transplantation. This survey is chosen because patients were asked a three-item cognitive ability test after completing the choice tasks and hence it is a highly relevant survey to explore the effect of cognitive ability on response consistency in CEs. The result shows that higher cognitive ability is positively associated with scale and hence a lower error variance. Patients who answered all the three-item cognitive ability test correctly can state consistent choices, and these group of patients have a lower WTW for changes in attributes of kidney transplantation compared to those who answered the three-item questions imperfectly.

The remainder of the paper is organised as follows. In the next section, I summarise the literature on the role of cognitive ability in financial, economic, and medical decisions both in revealed and stated preference context. The literature provides an insight into why cognitive ability is important in stated preference experiments. Section 3.2 describes the experimental setting. In the next section, the modelling approach to investigate the role of cognitive ability on the variance of the error term is developed. Following that, the results and a general discussion, and an overall conclusion are presented.

3.1.1 Literature review

An increasing number of studies indicated that variation in individuals' cognitive skills has implications for financial, economic, medical, and many other decisions. One primary reason why cognitive ability might matter is that of the importance of information. Among studies in financial decision-making, for instance, Pak and Babiarz (2018) examined whether or not cognitive decline led to a safer portfolio choice and showed a strong positive correlation between cognition and stock ownership. Dohmen et al. (2010) investigated the link between cognitive ability, risk aversion, and impatience, using a choice experiment that measured risk aversion and impatience over an annual horizon, and found that lower cognitive ability is associated with larger risk aversion and more noticeable impatience. Christelis et al. (2010) studied the relation between cognitive abilities and stockholdings and found that the propensity to invest in stocks is heavily associated with cognitive skills.

Agarwal and Mazumder (2013) analysed the effects of cognitive abilities on consumer financial decisions and found that consumers with larger overall composite test scores, and particularly those with higher math scores, are less prone to make a financial mistake later in their life. The authors also found that the mathematical component of the test is what matters most for financial decision making and wealth. Vice versa, non-mathematical abilities appear to matter for non-financial forms of suboptimal behaviour (e.g. failure to take medicine). Banks and Oldfield (2007) showed that numeracy levels are strongly correlated with measures of retirement saving and investment portfolios, even after controlling for different dimensions of cognitive ability and educational achievement.

Other studies analysed the link between economic behaviour and cognitive abilities. Using a measure of intelligence called the Cognitive Reflection Test (CRT), which evaluates an individual's ability to override initial but incorrect cognitive impulses, Frederick (2005) showed that individuals with high CRT are more patient with short-term tradeoffs but not significantly more patient with long-term tradeoffs, suggesting that these differences may be due to variations in cognitive reflection rather than underlying preferences. Brañas-Garza et al. (2012) indicated that subjects with higher scores on the CRT test are more prone to play according to the Nash equilibrium in beauty contest game. Besides, individuals with lower cognitive ability are found to be more risk-averse (Burks et al. 2009; Benjamin et al. 2013). In small-stake lotteries, individuals with high standardised test scores are less likely to exhibit risk aversion (Benjamin et al., 2013).

Graffeo et al. (2015) investigated whether cognitive reflection and numeracy skills affect the quality of the consumers' decision-making process in a purchase decision context. Using retrospective verbal reports and eye-tracking experiments, the authors indicated that higher levels of cognitive reflection and numeracy skills predict the use of a more thorough decision process. They found that participants with a high CRT score chose the best deal more frequently, and showed a more thorough and detailed information search pattern compared to participants with a low CRT score.

In healthcare delivery, patients' engagement in shared decision-making is deemed vital to promote patient-centred care, and to that end, a growing number of studies within healthcare examined the effect of numeracy skills in medical decisions. Cokely et al. (2012) indicated that numeracy is the single strongest predictor of general decision-making skill, including risk literacy. Peters and Levin (2008) reported that people with higher numeric ability are less affected by whether outcomes are framed as gains or losses (framing effect). Cokely and Kelley (2009) suggested that individuals with higher numeracy make more consistent decisions.

In CEs, as to the best of my knowledge, only one study investigated the effect of cognitive functioning on the consistency of individual responses to a survey conducted exclusively with older people to examine preferences for multi-disciplinary rehabilitation (Milte et al., 2014). The authors showed that the presence of mild cognitive impairment (measured using the Mini-Mental State Examination (MMSE)) did not have a significant effect on the consistency of responses to the CE. However, the study was conducted only on a sample of older people (aged 65 years and older) and whether the result is consistent for another sample of individuals and setting is not clear. Moreover, the MMSE, a screening tool frequently used by healthcare providers to assess overall brain function, may not be the most relevant measure of cognitive ability.

A related study by LaRiviere et al. (2014) investigated how knowledge about a good affects willingness to pay (WTP) and scale, and how an objective signal (receiving one's score) causes changes in WTP and scale. The score was constructed from a short 8-multiple choice quiz on cold-water corals, and the participants' score was taken as a measure of knowledge about cold-water corals. The authors indicated that higher knowledge is associated with a more consistent choice process, but receiving an objective signal had no significant impact. The multiple quiz questions, however, were mainly focused on cold-water corals, which do not provide precise information about the cognitive ability of the respondents. In sum, the literature suggests that cognitive ability plays important role in many decisions, at least when comparing behaviour across different tasks in different fields.

3.2 The experiment

3.2.1 Context

This study used a choice experiment (CE) survey concerned with patients' time and risk preferences in kidney transplantation (Genie et al., 2018). Four attributes described each kidney transplant alternative: waiting time, expected graft survival, infectious risk, and neoplastic risk. These attributes were identified through a discussion with kidney surgeons. A detailed presentation of the attributes and their values are presented in Table 3.1.

Table 3.1: Attributes, description, levels, and coding scheme.

Attributes	Description	Levels	Coding (effect)
Waiting time	The number of months one has to wait to obtain the proposed transplant.	6, 12, 36, 60 (months)	Linear (-)
Graft survival	The expected length of time the kidney functions well.	10, 15, 20 (years)	Linear (+)
Infectious risk	The risk of contracting infectious disease by the transplanted organ.	Standard, Augmented	Dummy (+)
Neoplastic risk	The risk of contracting a tumour through the transplanted organ.	Standard, Augmented	Dummy (+)

Waiting time is the time that patients have to wait to obtain the proposed transplant. The waiting time depends on the characteristics of the recipient and the frequency with which donors of a particular type are available. This attribute allows patients' to evaluate an approximate waiting time, although there is a chance of waiting lower or higher than declared.

The expected graft survival attribute is determined by the features of the organ itself, the characteristics of the recipient and the compatibility between donor and recipient. This attribute allows respondents to understand how long the transplanted organ is functioning. The assessment is the result of a probabilistic calculation based on previous clinical data and experience of the doctor performing the evaluation, but it is subject to a certain degree of uncertainty.

Each risk attribute has two qualitative levels (standard and augmented risks). The wording chosen to describe the levels is the same as those used by the surgeons to describe marginal kidneys to their patients. Standard risk includes cases for which the evaluation process did not identify any risk factor for transmittable disease, and it is the most frequent condition in the assessment of donors and grafts. It is commonly defined as a standard risk to make clear to patients that a zero risk kidney does not exist since infectious or neoplastic pathologies can still be transmitted even if guidelines and good clinical practice are followed. In augmented risk, some of the controls have not been performed, or the donor had some risky behaviours or neoplastic disease in the days before his or her death, but an infection or neoplasm may still not result from clinical diagnostics (even if it is possible).

CEs commonly include a cost attribute to be able to compute WTP for changes in the composition of a good or service. However, this study did not include a cost attribute as its inclusion is unrealistic within a publicly provided health care system, as in Italy because medical services are typically free at the point of delivery and patients do not have an experience of paying for medical services.

3.2.2 Recruitment and ethics

The participants were 250, nearly the entire population of patients with end-stage renal disease (ESRD) enrolled on the waiting list for kidney transplantation at the Kidney and Pancreas Transplantation Unit, University of Padua, Italy. Personal characteristics of patients across the four numeracy scores are shown in Table 3.2. Only gender significantly differed across the scores. The questionnaire was administered, face-to-face, conducted by a trained interviewer. Ethical approval for the study was obtained from the Ethical Committee of Padua Hospital.

Table 3.2: Patient characteristics and numeracy score (NS)

	total N=248	NS				
		0	1	2	3	
		N=23	N=53	N=109	N=63	P-value[a]
Demographics						
Age	49.9 (11.3)	47.8 (13.8)	49.5 (11.4)	51.3 (10.6)	48.6 (11.3)	0.418
Time on dialysis	3.38 (3.76)	5.39 (5.79)	3.63 (3.86)	3.36 (3.64)	2.48 (2.52)	0.344
Gender						<0.01
Female	86 (35)	12 (13.95)	30 (34.88)	31 (36.05)	13 (15.12)	
Male	162 (65)	11 (6.79)	23 (14.2)	78 (48.15)	50 (30.86)	
Education						0.154[b]
Elementary	90 (36.29)	14 (15.56)	22 (24.44)	35 (38.89)	19 (21.11)	
High school	118 (47.58)	6 (5.08)	25 (21.19)	56 (47.46)	31 (26.27)	
College	40 (16.13)	3 (7.50)	6 (15.00)	18 (45.00)	13 (32.50)	
Marital status						0.193
Married	165 (66.53)	12 (7.27)	40 (24.24)	74 (44.85)	39 (23.64)	
Not married	83 (33.47)	11 (13.25)	13 (15.66)	35 (42.17)	24 (28.92)	
Employment status						0.05
Employed	155 (62.50)	12 (7.74)	27 (17.42)	69 (44.52)	47 (30.32)	
Unemployed	93 (37.50)	11 (11.83)	26 (27.96)	40 (43.01)	16 (17.20)	
Dialysis modality						0.734
Haemodialysis	181 (74.79)	18 (9.94)	38 (20.99)	76 (41.99)	49 (27.07)	
Peritoneal dialysis	61 (25.21)	5 (8.20)	13 (21.31)	30 (49.18)	13 (21.31)	

Continuous variables are presented as means with standard deviations. Categorical variables are expressed as frequencies and percentages. Numeracy score represents the total number of correct answers on the three-item test. [a]Chi-square test; and [b]Fisher's exact test.

3.2.3 Experimental design and the choice questionnaire

Experimental design is the combination of the attribute levels used to construct the alternatives included in the choice sets. The natural choice would be a full factorial design, namely an experiment that contains all possible combinations of the levels of the attributes. In this study, given the number of attributes and levels, a full factorial design gives rise to 48 possible scenarios ($2^2 * 3 * 4$) that can be combined into 1128 possible choices. Running a CE where respondents are asked to choose from 1128 choices is unfeasible. The standard practice to reduce the dimensionality of the experimental design is to pick a statistical efficiency measure and select a possible subset of choices that maximise such a criterion. This approach aims to optimise the informativeness of the selected choices, but have no connection with utility theory.

McFadden and Train (2000) showed that to link individual parameters' estimates obtained from a CE to individual preferences, a necessary condition is to assume preferences are complete, monotone and transitive. As a result, following Battiston et al. (2016), these assumptions are made at the design stage. An algorithm was employed that searches for a list of choice sets in which dominant alternatives do not appear, choice sets are not repeated, and the number of choice sets for which the answer can be inferred from the previous one is minimised (assuming transitivity and monotonicity).

The D-optimality criterion is the most popular optimality criterion to evaluate the design of choice experiments (Kessels et al., 2006), and the minimisation of the D-error, which is the inverse of the determinant of the Fisher information matrix, gives D-efficient designs. The D-error is given as:

$$D - error = (det(\Omega))^{1/K}$$

where Ω is the covariance matrix of the β_m's obtained from the logit model, and K is the number of parameters or the size of the matrix used as a scaling factor for the efficiency measure. A low D-error indicates a more efficient design (Bliemer and Rose, 2005a). The 'AlgDesign' package in R (Wheeler 2006; Aizaki and Nishimura 2008) was used to generate a D-efficient design of 16 choice sets with D-error of 0.23.

The experiment was carefully designed to be consistent with economic theory, allowing to obtain reliable measures of WTW for changes in transplant attributes. In each choice task, patients were asked to select their preferred alternative among two kidney transplantation options: which of the two treatments would you prefer? (See Table 3.3 for an illustration of the choice task format). Since completeness, transitivity and monotonicity were assumed at the design stage, an indifference (opt-out) option was not included among the possible answers to the choice sets.

Table 3.3: llustration of a choice task (Original in Italian)

	Treatment A	**Treatment B**
Waiting Time	6 Months	6 Months
Expected Graft Survival	20 Years	15 Years
Infectious Risk	Standard	Standard
Neoplastic Risk	Augmented	Standard
Your Choice ?	☐	☐

In addition to the CE, which is the main part, the questionnaire also included questions about the socioeconomic characteristic of the patients and validated numeracy questions used to measure patients' cognitive ability. The numeracy questions were taken from the Survey on Health, Ageing and Retirement in Europe (SHARE) and has been used in different decision-making such as in financial (Mazzonna and Peracchi 2012; Christelis et al. 2010) and economic decisions (Banks and Oldfield, 2007). These numeracy questions are also used in the English Longitudinal Study of Ageing (ELSA) (Banks and Oldfield, 2007). The test aims to establish the respondents' ability to do simple mathematical tasks, by asking them to carry out calculations based on real-life situations (Mazzonna and Peracchi, 2012). The SHARE numeracy questions were translated into the Italian language. The English version of the questions is given below.

3.2.4 The SHARE numeracy test

To measure respondents' cognitive ability, I used the SHARE[2] numeracy test. The SHARE numeracy test consists of a few questions involving simple arithmetical calculations based on real-life situations. The set of questions asked in the SHARE numeracy test are:

1. The probability of contracting an illness is 10 percent, how many people out of one thousand would be expected to get the disease?
2. In a sale, a shop is selling all items at half price. Before the sale the sofa costs 300 Euros. How much will it cost in the sale?
3. A second-hand car dealer is selling a car for 6,000 Euro. This is two-thirds of what it costs new. How much did the car cost new?

[2]For a test of cognitive ability to be useful; it must be short and simple. The SHARE numeracy test is a 3-item test that can be concluded in less than five minutes, and it is a good predictor of cognitive abilities, especially regarding mathematical abilities.

The number of correct answers[3] to these questions (zero to three) had been shown to be an essential indicator of cognitive ability (Frederick, 2005). If a patient does not give the correct answer for any of the questions, the score will be zero (0/3). The maximum attainable score was three correct answers. The distribution of the numeracy scores is shown in Figure 3.1. The mean score is 1.85. The majority had scored 2/3 (44%), about 22% has scored 1/3, and the remaining 9% and 25% of them scored 0/3 and 3/3, respectively.

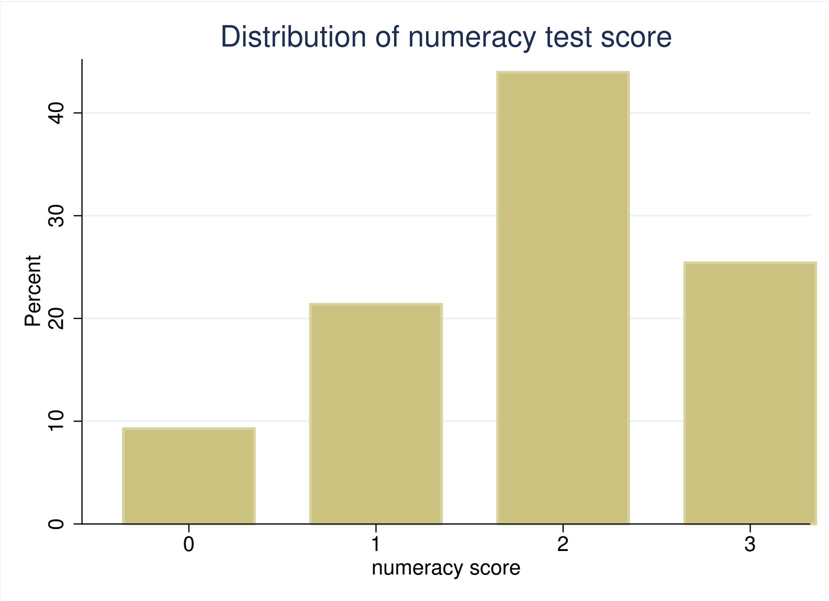


Figure 3.1: Distribution of numeracy test score

3.3 Incorporating cognitive ability in choice modelling

The analysis of responses obtained from CE is based on the random utility maximisation (RUM) framework (McFadden, 1974). The RUM framework consists of a collection of assumptions about the nature of individuals' choice behaviour. The three main behavioural assumptions include: (i) random utility (Thurstone, 1927): either the individuals do not know for sure the determinants of their decisions or the analyst cannot fully observe how participants make their decisions. In both cases, the utility attached to a good or service has a systematic (observable) component and a stochastic (unobservable) component. The stochastic component plays the role of modelling errors; (ii) multi-attributes utility (Lancaster, 1966): what matters to the individuals is not the utility of the good in itself, but the benefits given by its features. Then the systematic component of the utility of

[3]The correct answers for the questions 1-3 are 100, 150, and 9000, respectively. Throughout this paper, numeracy score and cognitive ability are interchangeably used.

a good is described as a combination of different attributes; and (iii) utility maximisation (Samuelson 1938; Manski 1977): the individuals act rationally and always select the good associated with the highest level of utility. This last behavioural assumption allows interpreting the respondents' choices as the results of a comparative process between the goods on offer (if option A is selected while options B and C were also available, this means that participants do prefer A over B and C).

According to the RUM, the utility of each kidney transplant alternative can be decomposed into a systematic utility (V) and a random utility (ε).

$$U_{ntj} = V_{ntj} + \varepsilon_{ntj} \tag{3.1}$$

where patients are indexed by $n = (1, ..., N)$, alternatives by $j = (1, ..., I, ..., J)$, i denotes the chosen alternative, choice tasks by $t = (1, ..., T)$, and attributes by $k = (1, ..., K)$. ε_{ntj} is an idiosyncratic error term assumed to be identically and independently distributed as type 1 extreme value (McFadden, 1974), which gives the multinomial logit (MNL) model. The indirect utility function (V) is typically assumed to be linear and additive:

$$V_{ntj} = \sum_k \beta_k X_{ntjk} \tag{3.2}$$

The indirect utility function accounting for cognitive ability or numeracy score (NS) is:

$$\begin{aligned} V_{ntj} = exp(\theta NS_{ntj})[\delta ASC_{ntj} + \beta_1 Waiting\ time_{ntj} + \beta_2 Survival_{ntj} + \\ \beta_3 Infectious\ risk_{ntj} + \beta_4 Neoplastic\ risk_{ntj}] \end{aligned} \tag{3.3}$$

where NS represents numeracy score and θ indicates the influence of numeracy on the error variance. The presence of the random component of the utility function means that the patients' choice behaviour becomes probabilistic and hence the probabilities of making a particular choice (P) is analysed as:

$$y_{nt} = i,\ if\ U_{nti} > U_{ntj}, \forall i \neq j \tag{3.4}$$

$$P(y_{nt} = i) = \frac{exp(\mu V_{nti})}{\sum_j exp(\mu V_{ntj})}, \quad and \qquad \mu = \frac{\phi}{\sqrt{6\sigma_\varepsilon^2}} \tag{3.5}$$

where μ is a scale parameter inversely related to the error variance usually normalised to unity as it cannot be identified independently from the error variance of the data σ_ε^2, imposing thus a constraint of constant errors variance (i.e., homoscedasticity[4]). However, this constraint can be relaxed by specifying the scale parameter as a function of observed variables, leading to a heteroskedastic multinomial logit (HMNL) model[5] (Bech et al. 2007; DeShazo and Fermo 2002; Hole et al. 2006). See Wright et al. (2018) for a systematic review of studies that accounted for scale heterogeneity in healthcare related choice experiments. In the HMNL model, the scale parameter μ is no longer constant as it allows for unequal variance across personal characteristics (survey characteristics). Using the HMNL model, the source of error variance such as cognitive ability (numeracy score) can be tested. The effects of cognitive ability and other potential factors on the error variance are modelled through a scale function:

$$\mu_n = exp(Z_n\theta) \tag{3.6}$$

$$P(y_{nt} = i) = \frac{exp(\mu_n V_{nti}\beta)}{\sum_j exp(\mu_n V_{ntj}\beta)} = \frac{exp(exp(Z_n\theta))V_{nti}\beta)}{\sum_j exp(exp(Z_n\theta))V_{ntj}\beta)} \tag{3.7}$$

where Z_n is a vector of individual characteristics including patients' cognitive ability and θ is a vector of parameters indicating the influence of the characteristics on the error variance. β and θ refer the effects on mean utility and scale, respectively. The estimates in the scale function tell us that whether a variable has a negative or positive effect on the error variance. It is also possible to assess the magnitude of the impact and compare the sizes across the variables included in the scale function (Bech et al., 2011). The HMNL[6] model collapses to the multinomial logit model when $\theta = 0$, and a test of $\theta = 0$ is a test for the error variance being constant across respondents (Hole et al., 2006).

Further, the generalised multinomial logit (GMNL) model[7] was employed to account for the panel nature of the choice data, that is, each patient provided repeated observations. The GMNL model also accounts for potential heterogeneity in the preference parameters and the heterogeneity of patients' errors (error variance) (Fiebig et al., 2010). In the GMNL model, following Fiebig et al. (2010), the utility to patient n from choosing a

[4]I explicitly model the effect of cognitive ability (numeracy score) on choice reliability through parametrisation of the scale factor.

[5]The HMNL is often used interchangeably with 'heteroskedastic conditional logit'. In this case, I use the term 'heteroskedastic multinomial logit' or 'HMNL' but recognise estimations in Stata for this definition will use the 'clogithet' command.

[6]The parameter vectors are estimated using maximum likelihood estimations (MLE) and using standard STATA routines. For this study, I used 'clogit', and 'clogithet' (Hole, 2009) procedures in STATA to estimate the MNL and HMNL models, respectively.

[7]See Fiebig et al. (2010) for the details of model derivations and estimation issues.

kidney transplant alternative j on choice set task t is given as:

$$U_{ntj} = \sigma_n ASC + [\sigma_n \beta + \gamma \eta_n + (1-\gamma)\sigma_n \eta_n] X_{ntj} + \varepsilon_{ntj} \tag{3.8}$$

where η_n is a vector of patient-specific deviations from the vector of mean preference parameter β; σ_n is a vector of patient-specific scale of the error term, which captures scale heterogeneity; γ is a parameter between 0 and 1, indicating whether the GMNL is constrained to GMNL-I or GMNL-II (Fiebig et al., 2010). The marginal utility for the k^{th} attribute can be represented as:

$$\beta_{nk} = \sigma_n \overline{\beta}_k + \gamma \eta_{nk} + (1-\gamma)\sigma_n \eta_{nk} \tag{3.9}$$

The elements of β_{nk} may differ from the sample mean $\overline{\beta}_k$ by η_n, which is a random variable with zero mean and standard deviation to be estimated. η_n helps to account for random heterogeneity in preferences, i.e. $Var(\eta_n)$. σ_n is a respondent-specific scale factor that shifts the whole vector of preference weights up or down in magnitude compared to the error term in the utility function (3.1). It is inversely correlated to the variance of that error term. When γ equals one, equation 3.9 becomes:

$$\beta_{nk} = \sigma_n \overline{\beta}_k + \eta_{nk} \tag{3.10}$$

Equation 3.10 is called GMNL-I (Fiebig et al., 2010), and the scaling σ_n affects only the means of the utility parameters. When γ takes the value of zero, equation 3.9 will collapse to:

$$\beta_{nk} = \sigma_n (\overline{\beta}_k + \eta_{nk}) \tag{3.11}$$

Equation 3.11 is called the GMNL-II model, suggesting that the scale affects both the mean and standard deviation of the parameters (Fiebig et al., 2010). GMNL-II assumes that the means and standard deviations vary proportionally with the scale factor. In GMNL-II the scale effect impacts both scale and taste heterogeneity.

Although different assumptions regarding the distributional form of η_n are possible, many applications use the normal distribution, i.e. $\eta_n \sim N(0, \sigma)$. Constraining neither the variance of η_n nor the scale factor σ_n defines the GMNL.

The probability that patient n chooses alternative i at choice task t in the GMNL framework can be presented as:

$$P(y_{nt} = i) = \frac{exp((\sigma_n\beta + \gamma\eta_n + (1-\gamma)\sigma_n\eta_n))X_{nti}}{\sum_j exp((\sigma_n\beta + \gamma\eta_n + (1-\gamma)\sigma_n\eta_n))X_{ntj}} \tag{3.12}$$

To identify which individual-specific characteristics drive the scale factor, σ_n can be parameterised as:

$$\sigma_n = exp(\overline{\sigma} + \theta' Z_n + \tau\nu_n) \tag{3.13}$$

$\overline{\sigma}$ represents a mean parameter of scale variance, $\theta' Z_n$ constitutes the systematic component of scale variation consisting of a vector of individual-specific variables Z_n (such as patients' cognitive ability) and an associated coefficient vector θ', τ is a parameter of unobserved scale heterogeneity and ν_n represents the unobserved scale heterogeneity (with $\nu_n \sim N(0,1)$). As σ_n must not change sign but shifts the whole coefficient vector up or down, it is defined as the exponential function of a normalising constant $\overline{\sigma}$, a systematic component $\theta' Z_n$ and a random part $\tau\nu_n$. Using 1,000 Halton draws to simulate the likelihood, the GMNL model with unconstrained γ parameter was estimated using 'gmnl' routines in Stata (Gu et al., 2013).

3.4 Results

Before using the numeracy score as a covariate in the scale function of the heteroskedastic model, an ordinary least squares (OLS) regression of the numeracy score on a set of variables was estimated to explore their effect on the respondents' cognitive ability formally. First, a summary of the results of an OLS regression of numeracy score on a set of factors that could potentially affect cognitive ability is presented. Following this, the model estimation results are presented.

3.4.1 Determinants of cognitive ability

Cognitive ability is not exogenous at the patient level, and hence casualty cannot be claimed from the estimates. Since cognitive ability (numeracy score) is itself a function of specific respondent characteristics, I estimated OLS regression model[8] with numeracy score as the dependent variable and a set of indicators as the explanatory variables. The

[8] An ordered logistic regression of numeracy score on a set of covariates was also estimated, and the results were reasonably consistent concerning the signs and significance of the included variables. Results are available upon request.

results are presented in Table 3.4. The analysis indicated that numeracy score was associated with the time spent on dialysis. Controlling for other factors, a longer time spent on dialysis has a negative effect on cognitive ability. Numeracy score was significantly affected by gender. Women had significantly lower numeracy score than their men counterparts. Frederick (2005) also found that women performed worse in CRT score compared to men. Numeracy score was not associated with patients' age and dialysis modality. The results show that college-level education does not significantly influence the numeracy score. As such, such a measure of cognitive ability should be independent of education level.

Table 3.4: Factors associated with numeracy score using an ordinary least squares (OLS) regression

	(1) (Coefficient)	(2) (Robust S.E.)
Dependent variable (numeracy score)		
Age	0.021	0.040
Age squared	0.000	0.000
Gender (female)	-0.502***	0.131
Education (college)	0.114	0.157
Marital status (married)	-0.031	0.126
Time in dialysis (years)	-0.043***	0.015
Employment status (working)	0.230*	0.136
Dialysis modality (haemodialysis)	0.136	0.128
_Cons	1.511	0.961
Number of respondents	242	
R-squared	0.1406	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The valid data used for the analysis was based on 248 patients. Of these patients, six went through pre-emptive kidney transplantation. Hence, information about dialysis modality for these patients was missing.

3.4.2 MNL model results-not accounting for scale differences

The MNL model (based on homoskedasticity of the error variance) serves as a reference for comparison. The results are presented in column 1 of Table 3.5. The result showed that all the attribute coefficients are statistically significant. The signs of the estimated attribute coefficients are as expected. From the MNL model we see, as expected, that the waiting time coefficient is negative and significant, indicating that, all else equal, patients are more likely to choose a transplant option with a shorter waiting time compared to an option with a longer waiting time. The marginal utility parameters for the two qualitative risk attributes are positive and significant, implying that the patients, on average, prefer a kidney transplant characterised by standard compared to augmented infectious and neoplastic risks. Similarly, the marginal utility for an organ that provides an extra year of survival is found to be positive and significant. The alternative specific constant-whose coefficient can be interpreted as a left-right bias in the choice task is positive and significant indicating the presence of left-right bias when completing choice tasks.

Table 3.5: The effects of cognitive ability on choice consistency: numeracy score the only covariate in the scale function

	(1) MNL Estimates	S.E.	(2) HMNL Estimates	S.E.	WTW	S.E.
Attributes						
Waiting time (β_1)	-0.033***	0.002	-0.030 ***	0.002	NA	NA
Graft survival (β_2)	0.185***	0.019	0.164***	0.019	5.381***	0.407
Standard infectious risk (β_3)	0.941***	0.058	0.850***	0.064	27.944***	1.337
Standard neoplastic risk (β_4)	0.995***	0.078	0.889***	0.082	29.237***	1.912
ASC (δ)	0.092 ***	0.035	0.087***	0.031	NA	NA
Covariates of scale						
Numeracy score (NS)			0.338 ***	0.115	-	-
Model statistics						
Number of observations	7936		7936			
Sample size	248		248			
Log-likelihood	-2415.374		-2411.3895			
BIC	4875.451		4876.423			

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The number of observations is obtained as (248 patients x 16 choice tasks x 2 alternatives=7936 observations). SE: standard error; BIC: Bayesian Information Criterion; ASC: Alternative Specific Constant. Numeracy score (NS) is a dummy variable taking a value of 1 if a patient answered all the numeracy questions correctly and zero otherwise.

3.4.3 Cognitive ability and scale

A binary variable 'NS' based on patients' score on the three-item cognitive ability questions was created. NS is an indicator variable which takes a value of one if the patient answered all the questions correctly (hereafter called 'high NS group') and zero otherwise (hereafter called 'low NS group'). A dummy numeracy score of two correct answers out of three questions (2/3) was also tested, but its effect was statistically insignificant. Patients who scored below three are those who incorrectly answered the third question. Thus, the decisive driver of scale, when it comes to cognitive ability, is answering the third question correct, which captured most of the effect on the scale factor. Numeracy score is included as a covariate of scale as explained in equation 3.6– to investigate–if higher numeraire patients (high NS group) are more predictable, on average, in their choices (Column 2 of Table 3.5).

Column 2 of Table 3.5 shows the results from using only numeracy score (coefficient of interest) as a covariate of scale. Patients with higher numeracy scores (being in high NS group) have a higher scale coefficient, i.e. the magnitude of the deterministic component of their utility functions relative to the random part is bigger than for patients with lower numeracy score (being in a low NS group). In other words, patients with a high numeracy score group are more consistent in their responses to the choice of the kidney transplantation alternatives.

Table 3.6: The effect of numeracy score (NS) on choice consistency across models: controlling for education, age, and gender

	(1) HMNL Estimates	S.E.	(2) GMNL Estimates	S.E.	SD	S.E.
Attributes						
Waiting time (β_1)	-0.025***	0.003	-0.050 ***	0.008	NA	NA
Graft survival (β_2)	0.125***	0.019	0.204***	0.038	0.244***	0.031
Standard infectious risk (β_3)	0.681***	0.074	1.371***	0.235	1.503***	0.152
Standard neoplastic risk (β_4)	0.708***	0.086	1.272***	0.229	1.273***	0.149
ASC (δ)	0.067 ***	0.025	0.152***	0.039	NA	NA
Covariates of scale						
Numeracy (NS) (score 3/3)	0.368***	0.121	0.992 ***	0.248	-	-
Education (college dummy)	0.326**	0.135	0.632 **	0.255	-	-
Age (+60 years)	0.140	0.139	0.553 **	0.243	-	-
Gender (female)	0.339***	0.116	0.580 ***	0.221	-	-
τ			1.173 ***	0.156	-	-
γ			0.551 ***	0.079	-	-
Model statistics						
Number of observations	7936		7936			
Sample size	248		248			
Log-likelihood	-2403.3182		-2115.5802			

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Numeracy score is a dummy variable equal to 1 if patients answered all the numeracy questions correctly, and zero otherwise. HMNL: heteroskedastic multinomial logit model; and GMNL: generalised multinomial logit model. In the GMNL estimation, graft survival, infectious risk, and neoplastic risk are assumed to be normally distributed, while waiting time and the alternative specific constant (ASC) are fixed. S.E: Standard Error, SD: Standard Deviation

In addition to the numeracy score, which is the primary variable of interest, other covariates that may affect cognitive ability and scale are included. The results are presented in Table 3.6. In the HMNL model, we see that the signs and significance of the marginal utility parameters are similar to those obtained under the basic MNL model (Table 3.5). Going through the parameters in the scale function, the variables numeracy score, college education, and gender (female) were all significant determinants of scale (Table 3.6). The effect of the numeracy score on the scale persists when the impacts of education, age, and gender are also accounted. Higher numeracy is associated with an increase in scale (reduction in error variance).

To account for the panel nature of the data, the analysis was repeated using the GMNL model. The signs and significance of the marginal utility parameters in the GMNL model remain the same as the other models. The SD parameter estimates in the GMNL model suggested a high level of preference heterogeneity across patients for each of the three attributes (graft survival, infectious risk, and neoplastic risk). Besides, the γ parameter appeared to be significantly different from 0, thus suggesting that the GMNL model could be collapsed into a GMNL-I model. The scale parameter τ is statistically significant, implying the presence of unobserved scale heterogeneity in the data considered in this study. In the GMNL model, allowing for preferences and scale heterogeneity did not noticeably change the structure of preferences for the attributes. While being older (60+ years of age) did not affect scale in the HMNL model, it had a statistically significant effect on the scale in the GMNL model.

Moving from standard MNL model to the HMNL and GMNL models, there are improvements in the model fit (Log-likelihood: -2415.374 vs -2403.3182 vs -2115.5802). In particular, for the GMNL model, the increase in model fitness could be partly explained by the fact that the panel nature of the choice data is accounted. Across the three models, bigger improvements are obtained when both preference and scale heterogeneity are considered (GMNL model). However, all the other conclusions remain the same across HMNL and GMNL models: mainly that the numeracy score is positive and statistically significant in affecting scale (Table 3.6).

3.4.4 Patients' cognitive ability and willingness to wait (WTW)

From a policymaking perspective, an important question is whether the WTW for changes in the multi-attribute content of the kidney transplant varies between respondents with lower and higher cognitive ability. To show how cognitive ability, through choice consistency, is correlated with willingness to wait (WTW) estimates, I estimated the MNL model in which numeracy score (NS) interacts with the attributes. The indirect utility function accounting for the interaction between NS and choice attributes is given as:

$$\begin{aligned}V_{ntj} = \delta ASC_{ntj} + \beta_1 Waiting\ time_{ntj} + \beta_2 Survival_{ntj} + \\ \beta_3 Infectious\ risk_{ntj} + \beta_4 Neoplastic\ risk_{ntj} + \gamma ASC_{ntj} * NS_n + \\ \beta_5 Waiting\ time_{ntj} * NS_n + \beta_6 Survival_{ntj} * NS_n + \beta_7 Infectious\ risk_{ntj} * NS_n + \\ \beta_8 Neoplastic\ risk_{ntj} * NS_n\end{aligned} \tag{3.14}$$

After estimating the MNL model, for instance, the WTW for an additional year of graft survival is computed as follows:

$$\frac{\partial V_{ntj}/\partial Survival_{ntj}}{\partial V_{ntj}/\partial Waiting\ time_{ntj}} = \frac{\beta_2 + (\beta_6 * NS_n)}{\beta_1 + (\beta_5 * NS_n)} \tag{3.15}$$

Table 3.7: The effects of numeracy score on taste parameters and WTW

	Main effects	**Interactions (NS)**	**WTW (months)**	
			Low NS	High NS
	Estimates	Estimates	Estimates	Estimates
Attributes				
Waiting time (β_1)	-0.030***	-0.015***	NA	NA
	(0.002)	(0.005)	-	-
Survival (β_2)	0.192***	-0.020	6.448***	3.812***
	(0.022)	(0.043)	(0.514)	(0.629)
Infectious risk (β_3)	0.913***	0.170	30.629***	24.004***
	(0.066)	(0.142)	(1.808)	(1.928)
Neoplastic risk (β_4)	1.000***	0.037	33.549***	22.993***
	(0.090)	(0.184)	(2.550)	(2.787)
ASC (δ)	0.083**	0.048	NA	NA
	(0.040)	(0.083)	-	-
Model statistics				
Number of observations	7936			
Sample size	248			
Log-likelihood	-2403.6421			
BIC	4896.69			

NS: numeracy score; BIC: Bayesian Information Criterion. Standard errors in parentheses ($^*p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$). Using a t-test (results not reported here), the differences in WTW between low and high NS groups are statistically significant.

In Table 3.7, in addition to the coefficients associated with each attribute (main effects), the interactions of each attribute-specific coefficient with a binary indicator of cognitive ability (NS) are provided. The labels 'low NS group' represents patients that scored imperfectly, while 'high NS group' represents patients that scored all the numeracy questions correctly. The results indicated that patients who provided three correct answers out of three questions (high NS group) are willing to wait less for a kidney transplant with better-expected outcomes (extra year of graft survival, standard infectious risk, and standard neoplastic risk) compared to low NS group. Patients with high NS had significantly higher marginal dis-utility associated with 'waiting time', and therefore are willing to wait significantly lower for each of the three attributes. Here, the effect of cognitive ability on WTW is through choice consistency. The WTW estimates are lower for consistent choices compared to inconsistent responses.

In summary, higher cognitive ability significantly increased scale (lower choice randomness) and the data based on the consistent choice resulted in a lower WTW estimate for changes in the multi-attribute content of kidney transplantation. The effect of cognitive ability on the WTW estimates is through choice consistency (that is, patients who were consistent in their responses had a lower WTW for changes in attributes of kidney transplantation). Patients with high cognitive ability make non-random decisions compared to the low numeracy score group; i.e., they respond more consistently than the low numeracy score group.

3.4.5 Robustness checks

In this section, I test the sensitivity of the main results. To check whether the results are robust to implementing numeracy as a continuous variable, I re-estimated the HMNL model replacing the dummy numeracy score (3/3) by a continuous numeracy score (0 to 3). That is, I treated numeracy score as a continuous variable in the scale function. The result indicated that higher score has a significant positive effect on the scale and hence patients with a higher numeracy score were more likely to make a consistent choice. The results are presented in Table 3.8 (columns 3 and 4).

Table 3.8: HMNL Models: robustness checks for numeracy score (dummy vs continuous)

	(1) HMNL-I	(2) HMNL-II	(3) HMNL-III	(4) HMNL-IV
Attributes				
Waiting time (β_1)	-0.0304***	-0.0246***	-0.0244***	-0.0178***
	(0.00215)	(0.00250)	(0.00345)	(0.00318)
Graft survival (β_2)	0.164***	0.125***	0.134***	0.0914***
	(0.0187)	(0.0187)	(0.0228)	(0.0195)
Standard infectious risk (β_3)	0.850***	0.681***	0.680***	0.493***
	(0.0643)	(0.0735)	(0.102)	(0.0907)
Standard neoplastic risk (β_4)	0.889***	0.708***	0.729***	0.526***
	(0.0820)	(0.0857)	(0.111)	(0.0987)
ASC (δ)	0.0865***	0.0668***	0.0658**	0.0455**
	(0.0315)	(0.0255)	(0.0268)	(0.0196)
Covariates of scale				
Numeracy (NS) (score 3/3)	0.338***	0.368***	-	-
	(0.115)	(0.121)	-	-
Age (60+ years)	-	0.140	-	0.141
	-	(0.139)	-	(0.139)
Gender (female)	-	0.339***	-	0.397***
	-	(0.116)	-	(0.122)
Education (college)	-	0.326**	-	0.309**
	-	(0.135)	-	(0.136)
NS (continuous)	-	-	0.163**	0.209***
	-	-	(0.0641)	(0.0688)
Model statistics				
Observations	7936	7936	7936	7936
Sample size	248	248	248	248
Log-likelihood	-2411.389	-2403.318	-2411.89	-2402.553
BIC	4876.423	4887.102	4877.424	4885.572

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

HMNL-I: HMNL model with dummy numeracy score of answering 3/3 as the only covariate of scale; HMNL-II: the same as HMNL-I but controlling for age, gender, and education; HMNL-III: HMNL model with continuous numeracy score as the only covariate of scale; and HMNL-IV: the same model as HMNL-III but controlling for age, gender, and education.

I also estimated different versions of the GMNL model (i.e., GMNL-I and GMNL-II models), where γ is constrained to be equal to zero and one (convergence is achieved more easily in these restricted models in STATA). These specifications avoid the confounding of random preference and scale heterogeneity by setting, for instance, $\gamma = 1$ so that the scaling only applies to the deterministic part of the preference weights (Equation 3.10), i.e. $\beta_{kn} = \sigma_n \overline{\beta} + \eta_{kn}$. The results are presented in Table 3.9.

The third and fifth column of Table 3.9 report GMNL models with $\gamma = 0$ and $\gamma = 1$, respectively. In general, patterns of coefficients are similar to the one in the HMNL and GMNL models, with some exceptions. The significance pattern of all other coefficients is the same as in the HMNL model. In all specifications of the GMNL model, the parameter τ is estimated. Note that γ is constrained to 1, which yields the GMNL-I model so that the scale parameter only applies to the coefficient means but not to their random components (Fiebig et al., 2010). I also ran models where γ was estimated freely or γ set to zero (the latter yields a model in which random scale is multiplied with random preference heterogeneity, which Hess and Rose (2012) argue cannot be separated), but these specifications did converge despite the criticisms. The statistically significant estimate of the scale heterogeneity parameter, τ, in all the GMNL models implies that there is unexplained (by covariates) scale heterogeneity in the population of patients.

Table 3.9: The effect of numeracy score (NS) on choice consistency across GMNL models: controlling for education, age, and gender

	(1) GMNL	(2) SD	(3) GMNL-II	(4) SD	(5) GMNL-I	(6) SD
Mean of coefficients						
Waiting time (β_1)	-0.0503***		-0.0553***		-0.0467***	
	(0.00825)		(0.00770)		(0.00665)	
Graft survival (β_2)	0.204***	0.244***	0.298***	0.274***	0.167***	0.267***
	(0.0378)	(0.0306)	(0.0474)	(0.0440)	(0.0304)	(0.0303)
Infectious risk (β_3)	1.371***	1.503***	1.570***	1.307***	0.780***	1.717***
	(0.235)	(0.152)	(0.250)	(0.196)	(0.173)	(0.127)
Neoplastic risk (β_4)	1.272***	1.273***	1.522***	1.226***	0.900***	1.497***
	(0.229)	(0.149)	(0.238)	(0.209)	(0.171)	(0.139)
ASC (δ)	0.152***		0.170***		0.104***	
	(0.0388)		(0.0446)		(0.0388)	
Covariates of scale						
NS	0.992***		0.718***		0.614***	
	(0.248)		(0.196)		(0.193)	
Age (+60 years)	0.553**		0.442**		0.458**	
	(0.243)		(0.191)		(0.223)	
Gender (female)	0.580***		0.332*		0.236	
	(0.221)		(0.173)		(0.188)	
Education (college)	0.632**		0.497**		0.558***	
	(0.255)		(0.211)		(0.208)	
τ	1.173***		0.838***		1.086***	
	(0.156)		(0.115)		(0.170)	
γ	0.551***		0		1	
	(0.0790)		(-)		(-)	
Model statistics						
Observations	7936		7936		7936	
Sample size	248		248		248	
Log-likelihood	-2115.5802		-2125.528		-2124.0159	
BIC	4356.329		4367.284		4364.26	

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

3.5 Discussion and conclusions

Choice experiments (CEs) are increasingly used in applied economics to measure individuals' preferences for the different aspects of non-market goods. In CEs, individuals face trade-offs while completing choice tasks and the consistency/reliability of choice responses have important implications for welfare analysis and computations of the marginal rate of substitution. To reliably use the responses for decision-making, patients need to provide a consistent (less random) answers to choice tasks. This paper set out to examine the effects of cognitive ability on choice consistency and its possible effect on the WTW estimates in a choice experiment. I made use of an existing dataset concerned with patients time and risk preferences for kidney transplantation in Italy.

There are principally two main results of this study. First, there is strong evidence that cognitive ability (numeracy score) has a positive effect on the scale. Econometrically speaking, a higher cognitive ability increases choice consistency in the sense that the relative weight of the deterministic part of the underlying indirect utility function increases relative to the error component. Hence, the analyst can make better predictions of choice probabilities. So, controlling for other factors, cognitive ability affected responses through a second channel (scale): the effect of cognitive ability on the scale is positive and statistically significant at 1% level in all the model specifications. Patients who answered all the three-item cognitive ability tests state, in effect, less random choices. Second, patients with high cognitive ability (data of more consistent choices) have a lower WTW for changes in the multi-attribute content of kidney transplantation.

In all the estimated models, high cognitive ability resulted into choice consistency over the valuation of the attributes of kidney transplantation. That variance decreased with cognitive ability could be explained by the understanding of the choice tasks, i.e., patients with high numeracy score may understand the choice tasks better and hence were more consistent in their choice responses. As a result, the consistency of the choices matters for stated WTW estimates. Higher impatience, given by a lower WTW, was more pronounced among individuals with high cognitive abilities. Here, the effect of cognitive ability on the WTW is through choice consistency. That is, patients with high cognitive skills tended to provide more consistent responses to the choice questions, and these group of patients have a lower WTW for changes in the kidney transplantation attributes compared to patients with low cognitive skills. The lower WTW among patients with higher cognitive ability may also be due to a higher opportunity cost of staying longer on dialysis. Individuals with a higher cognitive ability may have a higher opportunity cost of being on dialysis (waiting for a transplant), and hence they are more impatient. If responses are random, we must be careful on the value of the WTW estimates obtained from such CEs before using the result for cost-benefit analysis or any other decision-making.

The analysis revealed that only female gender (negatively) and the time spent on dialysis (negatively) influenced cognitive ability (numeracy score). Numeracy score was essentially uncorrelated with a college education, suggesting that poor numeracy score may also be prevalent even among those with a college education. This is consistent with a study by

Lipkus et al. (2001), who indicated that even highly educated individuals do not always comprehend numbers when making decisions. In this respect, CE practitioners should be aware of these flaws and avoid assuming that educated respondents possess the required skills to understand and consider each piece of information. The effect of cognitive ability is shown to persist after the effects of age, gender, and education are controlled. As a result, the impact of cognitive ability on the scale was coming from that variable *per se* rather than other determinants of cognitive ability.

Further, over 50% of the patients answered two or fewer of the three item numeracy questions correctly, suggesting the presence of lower numeracy skills in patients with end-stage renal disease patients. Only 25% of the patients answered the three questions correctly. Such prevalence of low numeracy score has implications in terms of understanding choice tasks and the consistency of responses to the choice questions.

Besides, from data quality and welfare analysis viewpoints, it may be tempting to 'clean' datasets from patients above a certain cognitive threshold. However, the framework used in the present study did not identify the level of the threshold needed to have a clean dataset. Instead, this analysis is intended to provide a framework to assess the link between the consistency of respondents answer to stated preference questions and cognitive ability.

Often, the objective in several stated preference experiments is to obtain a marginal rate of substitution (MRS) estimates or welfare measures; and since all these estimates may depend on the degree of choice randomness, an appropriate (more objective) instrument should be in place to check the consistency/reliability of responses before using the results for decision-making. In this regard, the cognitive ability of respondents could be taken into account when making assumptions about individual choice behaviour. Providing training to respondents about the good or service to be valued before conducting the actual experiment may help improve their understanding and obtain consistent responses potentially. Moreover, the result highlighted that we should be cautious in assuming constant random errors as they may not be constant.

Many CEs include non-experimental tasks to test monotonicity[9] and stability[10] of preferences. Besides, debriefing (supplementary) questions are included to measure individuals' understanding of choice tasks. Individuals may hold stable and monotonic preferences and provide subjective answers to the debriefing questions, but these do not necessar-

[9] Monotonicity refers to dominant (desirable) features are preferred to less of an undesirable feature (Krucien et al., 2017) and are included to check whether participants hold monotonic preferences when they are presented with choice tasks in which one alternative is more attractive than the other.

[10] Stability test implies the ability that participants make the same choices when they are given the same choice tasks (Krucien et al., 2017)

ily provide adequate information about the consistency of choice responses. Therefore, CE practitioners should include a few questions (that can quantify cognitive ability) to objectively measure the consistency of responses instead of solely relying on subjective debriefing questions and non-experimental choice tasks. Besides numeracy questions, future CE studies in health care may also incorporate other dimensions of cognitive ability to ascertain its possible effect on response consistency and reliability.

Although this study underlines the importance of incorporating a cognitive ability test to determine choice consistency in CEs, there are some limitations. Firstly, while I focus on cognitive ability, I recognise that other patient and survey characteristics may affect response consistency not addressed in this paper.

Secondly, there may be the risk of not accurately observing the cognitive ability of patients. In this paper, I used validated numeracy questions as a proxy for cognitive function. There may be a potential confounding effect between cognitive function and commitment as well as capability. A patient might have a high cognitive ability but not much concerned with answering the numeracy questions carefully due to a lack of commitment as well as capability. I recognise the difficulties in separating commitment and capability in answering the numeracy questions, that is, one cannot differentiate between the case where low cognitive function is the outcome of a lack of capability and where it is instead a lack of commitment (as they may think the questions trivial or odd).

Finally, I recognise that the numeracy questions may not be a perfect measure of cognitive ability. In this regard, it is also essential to consider other elements of cognitive function such as tests of orientation in time, memory and verbal fluency.

To sum up, this paper highlighted the importance of accounting for cognitive ability measures to objectively determine the consistency of responses obtained from choice experiments, suggesting the implications for WTW estimates.

Chapter 4

Attributes aggregation in multi-attribute choice: Does it exist?

ABSTRACT

Choice experiments (CEs) are commonly employed in economics to value non-marketed goods. Within CEs it is assumed that individuals consider all attributes and make a trade-off between them. However, attributes-based decision-making is cognitively demanding, often triggering the adoption of alternative decision rules. There is growing interest in heuristics that individuals use when processing multi-attribute choice information. We develop a new framework in which individuals restructure the multi-attribute information into a meta-attribute before making their decisions. We estimate a non-linear utility model allowing attribute aggregation (AA) to depend on the information structure. This new model assumes participants are more likely to aggregate information into a meta-attribute when the attributes provide similar information about the good or service. Accounting for AA leads to improvements in model fit. The probability of adopting this decision-making strategy is greater for homogenous information. AA is more prevalent amongst participants who adopted a quick and click strategy (shorter response time), more likely to occur for later positioned choice tasks (potentially due to fatigue effect), leads to improvements in model fit and has implications for welfare estimates. Our results underline the importance of accounting individuals' information processing rules when modelling multi-attribute choices. [1]

[1]This work was conducted while I was a visiting Ph.D. student at the Health Economics Research Unit (HERU), the University of Aberdeen (United Kingdom), from September 2017 to March 2018 (working with Dr. Nicolas Krucien and Professor Mandy Ryan). The paper is co-authored with Mandy Ryan and Nicolas Krucien.

4.1 Introduction

Choice experiments (CEs) are commonly used in applied economics to value non-market goods (Louviere et al., 2000). In a CE, participants are typically asked to choose between two or more multi-attribute hypothetical descriptions of the good. These stated preferences are used to estimate the marginal utility of changes in the composition of the good. This is made possible by Lancaster's theory of demand (LTD) which stipulates that utility of the good comes from its attributes rather than the good itself (Lancaster, 1966). When analysing individuals' choices, it is commonly assumed that individuals are willing to trade the attributes,[2] such that less of one attribute can always be compensated by more of another attribute. This assumption of compensatory choice behaviour is important, allowing estimation of marginal rates of substitution (MRS) between attributes (de Bekker-Grob et al. 2012; Lancsar and Louviere 2008; Ryan et al. 2007).

However, there is growing evidence in the CE literature that individuals can adopt a non-compensatory choice behaviour. For example, they can decide to ignore some attributes (Alemu et al. 2013; Erdem et al. 2015; Heidenreich et al. 2018; Hensher 2006; Hole 2011; Hole et al. 2013; Lagarde 2013), or eliminate/select choice alternatives based on some decision criteria (e.g., 'elimination-by-aspects') (Erdem et al. 2014; Gilbride and Allenby 2004; Tversky 1972). Such rules imply discontinuities in the demand function and preclude the computation of MRS or lead to biased results if not accounted for (Campbell et al. 2011; Heidenreich et al. 2018; Erdem et al. 2014).

Our study examines another type of discontinuity in demand function referred as attributes aggregation (AA). Under AA, individuals would restructure the multi-attribute information by combining several attributes into one new piece of information (hereafter meta-attribute). For example, when comparing different food options, instead of directly comparing the options in terms of fat, saturated fat, sugar, and salt, the individuals could first re-structure this nutritional information by combining all four attributes into a "product healthiness" dimension, and then consider this meta-attribute alongside other relevant attributes such as cost and quantity. By doing so, individuals effectively change the nature of the decision problem, comparing options in terms of healthiness rather than on every single nutritional attribute. This editing of the multi-attribute information has important consequences for the identification of the demand function. It implies that the combined attributes (e.g., salt) can no longer be traded against the non-combined attributes (e.g., cost), making thus the computation of MRS[3] impossible. Whilst AA does

[2]Technically this takes the form of a utility function additive in its arguments.

[3]AA does not describe a case of non-compensatory choice behaviour, as MRS is still possible between the non-combined attributes and the meta-attribute (i.e., new dimension arising from the attributes

not question the validity of LTD, as individuals' choice behaviour remain based on the product attributes, it interrogates how the LTD has been applied to identify preferences for multi-attribute goods. By assuming that product attributes would separately influence the decision-making, CEs might have taken the analysis of consumer demand from one extreme (where product attributes were irrelevant) to the other (where each product attribute would be allowed to influence the choices separately). However, multi-attribute information often has a structure (e.g., the different product nutrients being conceptually related) and this should be reflected in the measurement of preferences.

AA is likely to occur in multi-attributes choices where good or service include attributes on a thematic structure. Vatn and Bromley (1994) described a translation effect following which individuals would convert the "objective" multi-attribute information (e.g., a loaf of bread can be broken down into calories, taste, smell, structure, and texture) into a "subjective" unidimensional information (e.g., healthiness). Whenever the attributes of the good are expected to share some conceptual relationships (e.g., food nutrients being all indicators of healthiness), this translation effect becomes relevant and should be accounted for when modelling individuals' preferences. For example, in a CE about health professionals' preferences for working conditions, (Kolstad, 2011) included monetary and non-monetary attributes which would refer to the extrinsic and intrinsic motivations of the respondents. In another CE about preferences for end-of-life healthcare, Barbara et al. (2018) broke down the cost information into several attributes (i.e., additional treatment cost per case, additional insurance premium per year, additional out-of-pocket costs) and then when making their decisions some participants might recombine this information into one single monetary dimension.

Whilst AA is likely to be relevant in many CEs, the literature has paid almost no attention to it. In the only study to our knowledge, Layton and Hensher (2010) showed that up to 88% of individuals aggregated attributes when making commuting decisions. The results indicated that accounting for AA significantly impacted the MRS between the attributes.

Our study contributes to the literature by developing a general AA model which could virtually be applied to any situation where AA is relevant. Whilst Layton and Hensher (2010), only allowed attributes sharing the same format to be combined (e.g., *"Number of minutes in free-flow traffic"* and *"Number of minutes in slowed down traffic"* were combined into a *"total traffic time"*), our new model extends the principle of AA to attributes which don't necessary share the same format, such as qualitative attributes.

The rest of this paper is organised as follows. Section 4.2 describes the experimental setting and data. Section 4.3 describes the choice modelling approach, providing a bench-

aggregation).

mark model for comparison, and then describes the AA modelling, explaining the behavioural process. Section 4.4 presents the model results. Accounting for AA improves model fit, with the probability of adopting AA greater for homogenous information. AA is more prevalent amongst participants who adopted a quick and click strategy (shorter response time) and more likely to occur for later positioned choice tasks (potentially due to fatigue effect). Section 4.5 presents the implications of AA for the monetary valuation of service improvements. Accommodating AA behaviour leads to a reduction in the WTP estimates. Section 4.6 discusses the results. Section 4.7 makes concluding remarks.

4.2 Experimental design and sample

We used a CE survey concerned with preferences for personalisation of chronic pain self-management programmes (Burton et al., 2017). Each choice option was described by four qualitative attributes: providing personalised information (INFO); providing advices that match personal situation (SITU); putting an emphasis on personal values in living well (LIVE); and communication style (COMM). In addition, a quantitative cost attribute was included (COST). The attributes and their levels are shown in Table 4.1. We chose this CE for an investigation of AA because the four qualitative attributes are indicators of one same underlying dimension (i.e., personalisation of the service).

Table 4.1: Attributes and levels for the choice experiment

Attributes	**Attribute description**	**Attribute levels**
Information	Information about pain, the conditions that cause it, and the different ways there are of managing it.	Provides everyone with the same information (NEUTRAL) Provides information that is relevant to you (HIGH)
Situation	Things like where you live, who you live with, what resources you have, what you usually do for yourself and others, and how pain currently affects that.	Takes little account of your current situation (NEUTRAL) Makes suggestions that fit your current situation (HIGH)
Living well	Things that really matter to you, especially the kinds of things that you would like to achieve or to spend more time doing, and the kind of person that you want to be.	Seems to think that everyone wants to get the same from life (NEUTRAL) Works with you on what you want to get from life (HIGH)
Communication	The way that the support service might communicate with you.	Communicates with you in a neutral professional way (NEUTRAL) Communicates with you in a friendly and personal way (HIGH)
Cost		£5, £10, £15, £20

Twelve experimental choice tasks, each with three choice options, were generated using a D-efficient design[4] (Bliemer and Rose, 2005b). In each task, participants were asked to choose their most preferred choice option (Figure 4.1). In addition to the 12 experimental tasks, each participant was asked to answer two non-experimental tasks which were manually added: a warm-up task (Task #1) to familiarise participants with the format of the choice questions and a monotonicity test (Task #14) (i.e., a dominance task in which one alternative dominates other alternatives). The questionnaire also included socio-demographic questions.

The sample consisted of 517 members of a UK-based online panel managed by a research company (*"ResearchNow!"*). These panel members were recruited via email and online marketing and represented a diverse range of people in terms of socio-economic indicators and medical histories. Invitations were targeted to 16+ years old panel members diagnosed with chronic pain. The study was approved by the North of Scotland Research Ethics Service. Survey and person-level characteristics of respondents are described in Table 4.2.

Table 4.2: Survey and participant-level characteristics

	Variable code	N	%
Education	0 (=less/other)	297	57.4
	1(=University)	220	42.6
Choice task position	Block 1 (tasks 1-4)	175	33.3
	Block 2 (tasks 5-8)	342	33.3
	Block 3 (tasks 9-12)	342	33.3
Monotonicity	Pass	176	86.46
	Fail	341	13.54

[4]Priors were obtained from a quantitative pilot study of 120 individuals.

Figure 4.1: Illustration of a choice task used in the experiment

research now™

44237901_UK Chronic Pain sufferers - Main_Uk_W1 *Live*

Please compare the three support services (A, B and C) and then answer the question below by clicking on the button for the service you choose. Please assume that each support service will be provided once a week for six weeks:

Service A:	• Provides information that is relevant to you • Takes little account of your current situation • Seems to think that everyone wants to get the same from life • Communicates with you in a neutral professional way • Costs £5 per week for six weeks

Service B:	• Provides information that is relevant to you • Makes suggestions that fit your current situation • Works with you on what you want to get from life • Communicates with you in a friendly and personal way • Costs £20 per week for six weeks

Service C:	• Provides everyone with the same information • Takes little account of your current situation • Works with you on what you want to get from life • Communicates with you in a friendly and personal way • Costs £10 per week for six weeks

Which service would you like the most?

Service A ☐ Service B ☐ Service C ☐

4.3 Attributes aggregation in multi-attribute choices

4.3.1 Conceptual framework

Multi-attribute choices are typically modelled within the random utility maximisation (RUM) framework, assuming random utility (Thurstone, 1927), multi-attributes utility (Lancaster, 1966), and utility maximisation (Samuelson, 1938). The utility (U) of each choice option is decomposed into a systematic (V) and random (ε) component (Equation 4.1). Following the Lancasterian theory of demand (LTD), V is defined by the product attributes (X) and their respective marginal utilities (β) (Equation 4.2). The presence of the random component means that the participants' choice behaviour becomes probabilistic (P) (Equation 4.3) and hence we can only predict choice probabilities. When the random component is assumed to be identically and independently distributed as type 1 extreme value, the choice probabilities can be represented by the multinomial logit (MNL) model (McFadden 1974; Train 2009).

$$U_{ntj} = V_{ntj} + \varepsilon_{ntj} \tag{4.1}$$

$$V_{ntj} = \sum_k (\beta_k X_{ntjk}) \tag{4.2}$$

$$P(y_{nt} = 1) = \frac{exp(V_{ntj})}{\sum_j exp(V_{ntj})} \tag{4.3}$$

U_{ntj} is the utility for participant n from alternative j in choice task t; β_k is the parameter to be estimated for the kth attribute; X_{ntjk} is the measured attribute k. As a reference model, we estimated an MNL model without accounting for AA.

The AA can be represented within the RUM framework by editing the indirect utility function (V). However, this requires making a number of assumptions about the nature of the AA process.

Assumption # 1: Which attributes should be considered for aggregation?

We define AA on the basis of conceptual proximity. The four qualitative attributes are conceptually related, together describing the "level of service personalisation".

The set of qualitative attributes eligible for AA is given as:

$$X_{ntjk}^{AA} = \{INFO_{ntj}, SITU_{ntj}, LIVE_{ntj}, COMM_{ntj}\} \tag{4.4}$$

The COST attribute is not eligible for AA and hence considered in the set of non-aggregated attributes (hereafter attributes partitioning (AP)):

$$X_{ntjk}^{AP} = \{COST_{ntj}\} \tag{4.5}$$

Assumption # 2: What triggers aggregation?

A decision maker faces and seeks to resolve a choice conflict in which the individual must select a choice from some set of alternatives (products, brand or generally choice objects) (Shugan, 1980), and then derives satisfaction from the product. Once conceptually-related attributes are identified, the respondents will decide to aggregate them depending on whether the attributes provide similar versus conflicting information. For example, if two attributes describe a high level of personalisation and the remaining two attributes describe a low level of personalisation, then AA is unlikely to be relevant as the attributes provide opposite information about the service. However, if the four qualitative attributes describe a high (low) level of personalisation, then the respondents are more likely to combine them into one single piece of information.

When determining what triggers AA, the analyst needs first to specify how the homogeneity of multi-attribute information is evaluated. In our study, respondents are assumed to evaluate the multi-attribute information by calculating a ratio of features[5] (φ).

$$\varphi_{ntj} = \frac{Min(Count_{ntj}^{High}, Count_{ntj}^{Neutral})}{Max(Count_{ntj}^{High}, Count_{ntj}^{Neutral})} \tag{4.6}$$

This ratio is computed by counting the number of high personalisation and neutral personalisation values and then taking their ratio. The ratio is defined such that its value is null when the four qualitative attributes provide the same information (i.e., four neutral (or high) attributes) and reaches its maximum value of one when the attributes provide mixed information (i.e., two neutral and two high attributes).

Once the homogeneity of multi-attribute information is evaluated, the respondents will decide whether the information is homogeneous enough to justify AA. This is done by

[5] We used standard deviation (SD) of the attributes' levels as an alternative measure of information homogeneity/heterogeneity, but the corresponding AA-RUM model was associated with a lower level of statistical performance. The results are available upon request.

comparing the objective (φ) ratio to a personal/subjective (α) threshold. If $\varphi > \alpha$ individuals retain the initial information structure (i.e., AP), and if $\varphi < \alpha$ individuals proceed to AA.

$$U_{ntj} = \begin{cases} V^{AP}_{ntj} + \varepsilon_{ntj}, & \text{if } \varphi_{nt} > \alpha \\ V^{AA}_{ntj} + \varepsilon_{ntj}, & \text{if } \varphi_{nt} \leq \alpha \end{cases} \tag{4.7}$$

where V^{AP} and V^{AA} corresponds to the indirect utility function under attributes partitioning (AP) and attributes aggregation (AA) respectively. This initial formulation can be enriched by allowing the threshold value to differ across participants (i.e. $\alpha \to \alpha_n$). That is, individuals are allowed to have preference heterogeneity for the threshold parameter. Similar to Layton and Hensher (2010), we allowed this threshold to vary across participants and choice tasks by specifying an exponential distribution with mean $\frac{1}{\lambda}$ and density $g(x) = \lambda e^{-\lambda x}$.

$$\alpha \sim exp(\lambda) \tag{4.8}$$

$$Distribution : P(X \leq x) = 1 - exp(-\lambda x) \tag{4.9}$$

The utility function allows a probabilistic attributes aggregation or partitioning. The probability of attributes aggregation (P^{AA}) is specified as:

$$P^{AA}_{nt} = 1 - P^{AP}_{nt} = exp(-\lambda \varphi_{nt}) \tag{4.10}$$

Assumption # 3: How is multi-attribute information aggregated?

AA indicates that attributes are combined together into a single dimension. However, the analyst still needs to determine how the combination actually happens. In our study, AA takes the form of a binary classification of the information. If the majority of qualitative attributes take high values, then the new dimension (or meta-attribute) would be "positive" taking a value of one. Alternatively, if the majority of attributes take neutral values, then the new dimension would take a zero value.

$$METAntj = \begin{cases} 1, & \text{if } (INFO_{ntj} + SITU_{ntj} + LIVE_{ntj} + COMM_{ntj}) \geq 3 \\ 0, & otherwise \end{cases} \tag{4.11}$$

This aggregation rule gives the same importance to the four qualitative attributes in the aggregation process. This is consistent with the Dawes' rule following which individuals make choices by counting the number of positive/good features and selecting the option with the highest count (Dawes, 1979).

Assumption # 4: Where does attribute aggregation take place?
In principle, AA is likely to be guided by both the choice task features and respondents' personal characteristics. We define AA at the choice task-level. That is, respondents are assumed to re-structure the multi-attribute information for none or all of the options included in the task. This approach prevents respondents applying AA for one option and AP for another, as this would imply some forms of asymmetric comparisons (e.g., meta attribute vs INFO) which are behaviourally difficult to justify. By assuming that the AA process takes place at the task level implies that first individuals evaluate the homogeneity of the multi-attribute information for each option (A, B, C) separately, leading thus to three measures ($\varphi_A, \varphi_B, \varphi_C$), and then, based on these measures, would decide to adopt or not AA as a decision rule.

$$P_{nt}(AA) = P(\varphi_{nt(A)} < \alpha_n) * P(\varphi_{nt(B)} < \alpha_n) * P(\varphi_{nt(C)} < \alpha_n) \tag{4.12}$$

$$P_{nt}(AA) = exp(-\lambda\varphi_{nt}), where\ \varphi_{nt} = \sum_j \varphi_{nt(j)} \tag{4.13}$$

4.3.2 Econometric specification

Following this specification of AA in RUM context, we estimate the following choice model:

$$\begin{aligned} U_{ntj} = (\beta_1 INFO_{ntj} + \beta_2 SITU_{ntj} + \beta_3 LIVE_{ntj} + \beta_4 COMM_{ntj})(1 - P_{nt}^{AA}) + \\ \beta METAntj(P_{nt}^{AA}) + \delta_1 ASC1_{ntj} + \delta_2 ASC2_{ntj} + \gamma COST_{ntj} + \varepsilon_{ntj} \end{aligned} \tag{4.14}$$

where P^{AA} corresponds to the probability of aggregating the multi-attribute information. The indirect utility function under attributes partitioning (V^{AP}) is specified as:

$$\begin{aligned} V_{ntj}^{AP} = \delta_1 ASC1_{ntj} + \delta_2 ASC2_{ntj} + \beta_1 INFO_{ntj} + \beta_2 SITU_{ntj} + \\ \beta_3 LIVE_{ntj} + \beta_4 COMM_{ntj} + \gamma COST_{ntj} \end{aligned} \tag{4.15}$$

The indirect utility function under attributes aggregation (V^{AA}) is specified as:

$$V_{ntj}^{AA} = \delta_1 ASC1_{ntj} + \delta_2 ASC2_{ntj} + \beta METAntj + \gamma COST_{ntj} \tag{4.16}$$

where $'META'$ corresponds to the meta-attribute obtained from the aggregation of the four qualitative attributes (i.e., INFO, SITU, LIVE, COMM).

The overall utility (Equation 4.14) is the weighted average of the two decision rules, where the weights are given by the probability of aggregation or disaggregation.

This initial formulation of the AA-RUM model can be enriched by allowing the threshold value to differ across participants. We also estimate an AA-RUM model allowing the threshold value to depend on survey and person-level characteristics.

$$\lambda = exp(\mu Z_n) \tag{4.17}$$

where Z_n represents personal characteristics of the respondents including education, whether the respondent pass the monotonicity test, the location of the choice tasks in the questionnaire, and the response time (RT) (short vs long response time), and μ is a vector of parameters indicating the influence of the characteristics on the aggregation threshold.

The model parameters to estimate are $(\beta_{1:4}, \beta, \delta_{1:2}, \gamma, \mu)$. We use 1,000 Halton draws to simulate the log-likelihood function (Train, 2009) using MATLAB[6].

4.4 Results

The results for the reference MNL[7] model not allowing for AA are presented in columns 2 and 3 of Table 4.3. All coefficients are significant in the expected directions (i.e., a positive effect for improvement in the personalisation dimensions and a negative effect for a COST increase).

Results of the RUM-AA model are presented in columns 4 and 5. We obtain a similar pattern of preferences for the five attributes. Allowing for AA specification improves the modelling performance as indicated by the decrease in the value of the Bayesian Information Criterion (i.e., 11,820.6 vs. 11,790.6). The coefficient of the meta-attribute

[6] Codes available up on request.

[7] We also estimated an error component logit (ECL) model but failed to outperform the MNL model, which merely increased the estimation time. Using a log-likelihood ratio test (LR test: Deviance=2.2; dof=3; P-value=0.532), we find no evidence of differences in respondents' choices between the two models and hence we focus our analysis based on the MNL model.

"personalisation" is positive and significant, implying that participants prefer a higher improvement in the service personalisation. For the RUM-AA model[8], $\lambda = 1.040$ ($p < 0.01$), indicating that our model accommodates a fraction of participants who adopt AA as a choice decision rule. Whereas perhaps the majority preserved attributes separately, 21.5% (111) are found to have adopted an AA processing strategy[9].

Table 4.3: Results of multinomial Logit (MNL) models

	non-AA (MNL)		AA-MNL		MNL-AA (with covariates)	
	MLE	S.E.	MLE	S.E.	MLE	S.E.
Information (β_1)	0.676***	0.028	0.801***	0.059	1.000***	0.083
Situation (β_2)	0.959***	0.039	1.183***	0.080	1.604***	0.132
Living well (β_3)	0.786***	0.034	0.938***	0.063	1.341***	0.104
Communication (β_4)	0.249***	0.028	0.242***	0.052	0.455***	0.055
Cost (γ)	-0.053***	0.003	-0.048***	0.003	-0.052***	0.003
ASC2 (δ_1)	0.165***	0.035	0.121***	0.036	0.049	0.038
ASC3 (δ_2)	0.005	0.034	-0.005	0.035	-0.023	0.035
META (β)	-	-	0.975***	0.216	0.062	0.119
Extent of aggregation (λ)	-	-	1.040***	0.169	1.148***	0.301
Covariates						
Education (University)	-	-	-	-	0.016	0.028
Dominance test (pass)	-	-	-	-	0.084***	0.029
Task position 2 (tasks 5-8)	-	-	-	-	0.187***	0.059
Task position 3 (tasks 9-12)	-	-	-	-	-0.203***	0.044
Response time (shorter)	-	-	-	-	-0.953***	0.284
Response time (longer)	-	-	-	-	0.987***	0.345
Number of observations	6204		6204		6204	
Sample size (N)	517		517		517	
Number of parameters	7		9		15	
Log likelihood	-5879.7		-5856		-5331.6	
BIC	11820.6		11790.6		11594.1	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

BIC: Bayesian Information Criterion; AA = Attribute Aggregation; ASC=Alternative Specific Constant; S.E.=Standard Error; MLE=Maximum Likelihood Estimator

To shed some light on the relationship between information heterogeneity and the probability of AA versus AP, Fig. 4.2 displays the variation in the probability of AP versus AA when information heterogeneity changes, *ceteris paribus*. The graph indicates that

[8] The λ parameter describes the extent of attributes partitioning/aggregation - a λ close to zero implies AA, as λ gets larger the standard AP model is optimal.

[9] To compute the number of attribute aggregators in the sample, we followed the following steps: First, we compute the AA probability for each task and individual; second, we compute the average AA probability for each individual; and finally, an individual is classified as an "aggregator" when the average AA probability is greater or equal to 50%.

the probability of AP increases with information heterogeneity. The negatively slopped part of the plot indicates that the probability of AA declines as information heterogeneity increases. For a very low level of information heterogeneity (for instance, φ=0.1), there is a 0.9 probability of AA, suggesting that more homogenous information is likely to be aggregated.

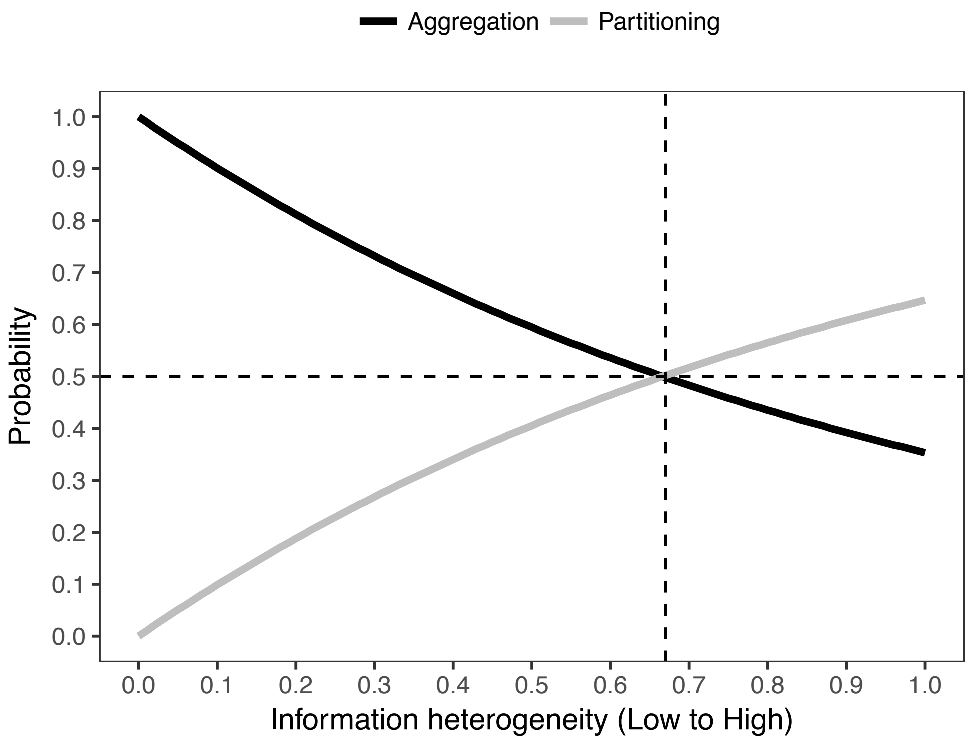


Figure 4.2: Cumulative density functions (CDFs) of attributes partitioning (AP) versus information heterogeneity

Accounting for heterogeneity in personal and survey characteristics further improves modelling performance, as indicated by the lower likelihood (LL) value ($LL_{AA_{heterogeneity}}$=5731.6 vs LL_{AA}=5856 vs LL_{no-AA}=5879.7). We find evidence of heterogeneity in the tendency to adopt AA rule (Table 4.3: columns 6 and 7). Controlling for other factors in the model, not passing the monotonicity (dominance) test has a negative effect on the aggregation threshold, with respondents who failed the monotonicity test having a lower threshold and hence more likely to adopt AA in choice decision-making. The sequence of the choice tasks in the questionnaire also affects the aggregation threshold. Compared to the first four choice tasks, respondents are less likely to adopt AA for the middle-positioned tasks (tasks 5-8) but are more likely to aggregate attributes when the choice tasks are located in the later positions (tasks 9-12). The time that respondents took to complete the choice tasks also affected the aggregation threshold. A shorter response time (RT) affected the threshold negatively, meaning respondents who spent a shorter time to complete the tasks are more likely to aggregate attributes compared to those who spent relatively a longer time. They are less likely to consider each attribute separately while completing choice tasks.

4.5 Implications of AA for the monetary valuation of service improvements/changes

From a policy perspective, an important question is whether accommodating AA affects individuals' WTP for changes in multi-attribute content of the service quality. We then compare WTP values between non-AA and AA models.

The general formula to compute WTP or the marginal rate of substitution (MRS) between an attribute (X_k) and COST is:

$$WTP_{X_k} = -\frac{\partial U_{ntj}/\partial X_k}{\partial U_{ntj}/\partial COST} \tag{4.18}$$

As COST was not allowed to be aggregated, the marginal utility of COST is the same in both non-AA and AA models: $\frac{\partial U_{ntj}}{\partial COST} = \gamma$

However, the marginal utility of the other (X_k) attribute differs between the two model specifications.

In the non-AA model:

$$\frac{\partial U_{ntj}}{\partial X_k} = \beta_k \tag{4.19}$$

In the AA model:

$$\frac{\partial U_{ntj}}{\partial X_k} = \beta_k(1 - e^{-\lambda\varphi_{nt}}) \tag{4.20}$$

$$\frac{\partial U_{ntj}}{\partial META} = \beta e^{-\lambda\varphi_{nt}} \tag{4.21}$$

To use the WTP formula for the AA model, we need to assign a specific value for φ. We considered five arbitrary[10] values between 0 and 1. Results are presented in Table 4.4 and Figure 4.3. Accommodating AA as a decision rule impacts on WTP. The AA model tends to generate lower WTP values, suggesting not allowing for AA in the analyses produces

[10] Any other values between zero and one can be possible. The values 0.1, 0.3, 0.5, 0.7 and 0.9 are chosen arbitrarily to check the changes in WTP as information heterogeneity changes. A value of 0.9 means that information is more heterogeneous compared to the other values.

biased WTP results. The average WTP for quality attributes derived from the non-AA model is biased upward compared to the AA model although the bias becomes very small for a higher levels of information heterogeneity (say, φ_{nt}=0.9). When information provided to the participants become less heterogenous (for instance, φ_{nt}=0.1), the differences in WTP values between the standard MNL and the AA-MNL models get bigger for all the personalisation attributes (Figure 4.4). Together with Fig. 4.3, these results show that there is a positive relationship between information heterogeneity (the x-axis) and the WTP for higher levels of each personalisation attribute.

Table 4.4: Willingness to pay (WTP, £) estimates for a higher levels of personalisation attributes

	(1)	(2)	(3)	(4)	(5)	(6)
			RUM-AA			
	non-AA	$\varphi = 0.1$	$\varphi = 0.3$	$\varphi = 0.5$	$\varphi = 0.7$	$\varphi = 0.9$
Information	12.755	1.648	4.473	6.766	8.630	10.143
Situation	18.094	2.434	6.606	9.993	12.745	14.980
Living well	14.830	1.930	5.238	7.924	10.106	11.878
Communication	4.698	0.498	1.351	2.044	2.607	3.064
Aggregated (META)	-	18.306	14.868	12.076	9.808	7.966

φ: the ratio of features measure indicating the extent of information homogeneity/heterogeneity. The values are arbitrarily chosen in the range of zero and one. φ=0.1: information is less heterogenous. φ=0.9: information is more heterogenous.

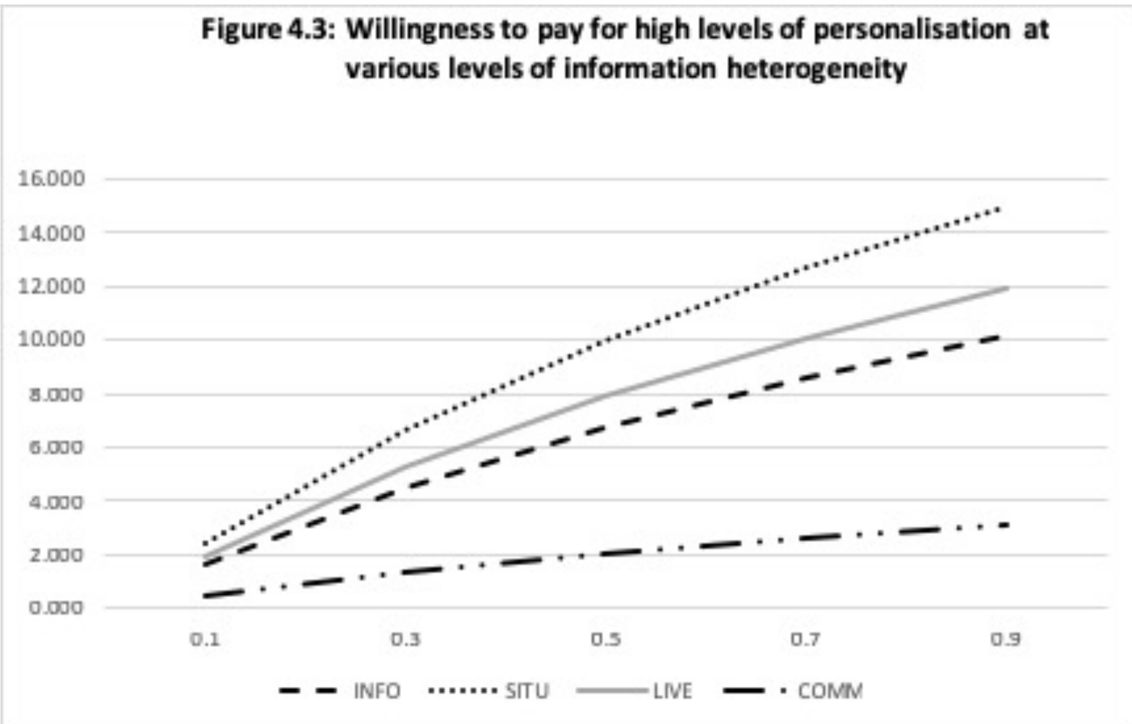


Figure 4.3: Willingness to pay for high levels of personalisation at various levels of information heterogeneity

Figure 4.3 shows the WTP for high levels of personalisation attributes at different levels of information heterogeneity. The X-axis represents the level of information heterogeneity and the Y-axis the WTP estimates: as heterogeneity increases (left to right), the WTP for high levels of each personalisation attribute declines. The valuation placed on the aggregated information is lower the more heterogenous information. Vice versa, individuals place a higher value to the aggregated information the more similar the given attributes.

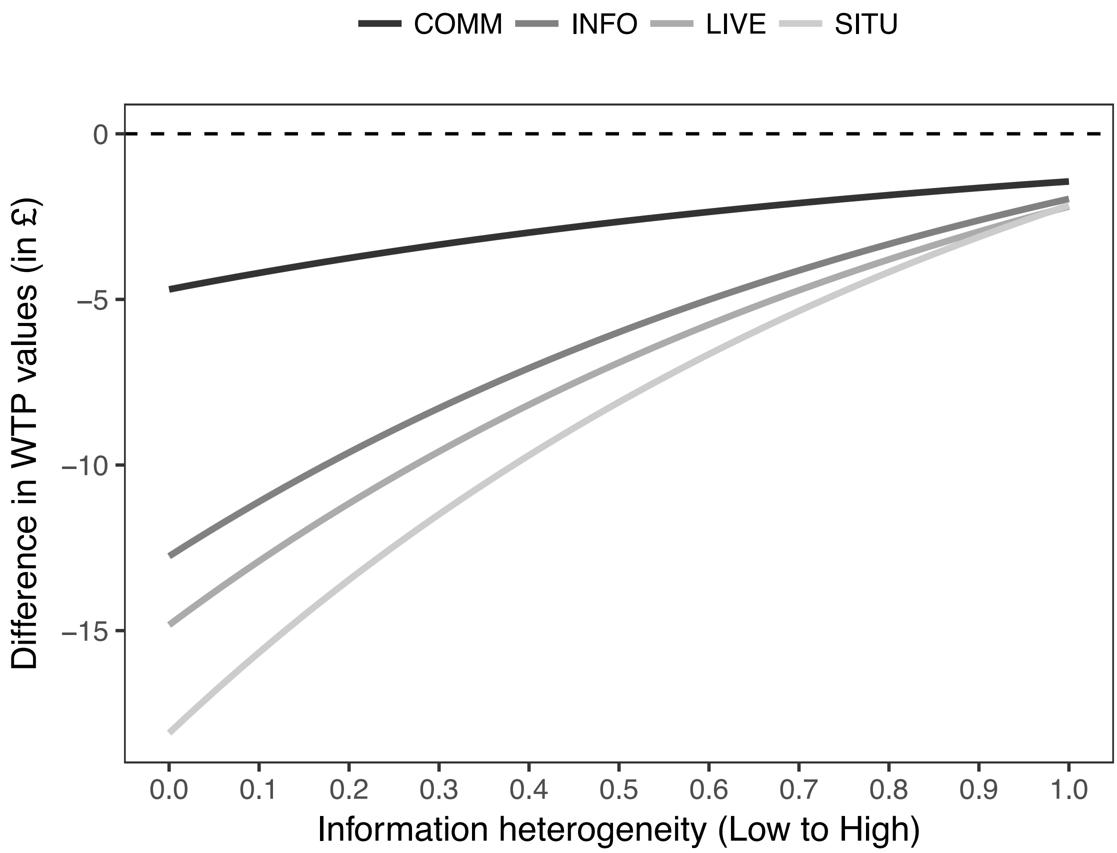


Figure 4.4: Differences in WTP values between standard MNL and AA-MNL models for different levels of information heterogeneity

The differences in WTP values between the standard MNL and AA-MNL models are indicated in Figure 4.4. The y-axis represents the difference in WTP between AA-MNL and the standard MNL models and the x-axis indicates the levels of information heterogeneity (heterogeneity increases from left to right). Accounting for AA results in a downward bias in the estimated WTP values. For a low level of information heterogeneity (say φ=0.1), the difference in WTP between the two models becomes larger and as the level of information heterogeneity increases (from left to right) this difference gets smaller and smaller. We see a smaller aggregation bias when information becomes heterogenous.

4.6 Discussion

Choice experiments are commonly employed in applied economics to value public goods. The standard model assumes respondents process each attribute separately. We develop a new framework in which a fraction of participants aggregates conceptually related attributes. We used a count measure, 'ratio of features', to assess information heterogeneity/homogeneity. We show that accommodating AA as a choice simplifying rule in stated preferences outperforms the standard full attributes partitioning model. Allowing for AA resulted in higher parameter estimates, impacting on willingness-to-pay estimates. AA is more prevalent among respondents who failed the monotonicity test, and those who adopted a quick and click strategy (faster RT). AA is also more likely to occur for the later positioned choice tasks.

In the only other study addressing attribute aggregation, Layton and Hensher (2010) showed that accounting for common-metric AA resulted in higher mean parameters and WTP estimates. However, our study indicated lower WTP estimates. One reason may be the difference in the methods employed to account for AA - whilst Layton and Hensher (2010) took a total score of common-metric attributes and the numerical distance between two common metric attributes as a means of evaluating information homogeneity, we maintained a binary classification of the attributes and used a ratio of features to assess information homogeneity. Another reason may be the difference in the CE setting. This suggests further investigation is required to understand AA better.

Whether AA should be explored in CEs depends on the nature or format of attributes. While aggregation of attributes based on a similar format as in the case of common-metric aggregation (Layton and Hensher, 2010) is one possibility, our approach is more general in the sense that other forms of attributes that do not share similar measurement units could be combined and evaluated together in multi-attribute choice. In this regard, our approach has some links with the literature on Hierarchical Information Integration (HII) with Integrated Choice Experiments (ICEs) (Molin and Timmermans 2009; Oppewal et al. 1994), which has been suggested to handle attributes in a complex decision-making task. The HII with ICEs assume that when individuals are confronted with complex decision problems, they initially divide sets of attributes that influence their choice behaviour into a smaller number, that is, into subsets (usually based on conceptual similarity or some form of thematic structure), then evaluate each subset separately and aggregate their evaluations of each subset to choose between competing opportunities (Louviere and Timmermans, 1990). HII suggests that individuals group similar attributes of choice alternatives into constructs.

In health economics, van Helvoort-Postulart et al. (2009) using HHI with choice experiment implemented a guideline for breast cancer surgery in day-care, a complex process involving changes at the organisational and management levels, as well as the level of health-care professionals and that of patients. In their study, conceptual proximity mattered when classifying attributes into different sets. For example, in a choice of implementation of breast cancer surgery in daycare, individuals may first separately evaluate attributes representing the 'organisation' set {day surgery unit, breast cancer nursing staff, compensation, discharge criteria, collaboration agreements with home care organisations}, 'cooperation partners' set {patients, colleagues, management, ward nurses, expertise home care}, and set of 'patient-centeredness' {written information after diagnosis, preoperative counselling, written information at discharge, possibility to choose between daycare and hospital admission, patient satisfaction}. Having made these separate evaluations, individuals integrate these separate evaluations to form preferences for competing breast cancer surgery options. The attributes in each set do not have the same measurement unit but were combined into constructs based on conceptual similarity. The AA framework adopted in our study is similar to each sub-experiment part of the HII integrated with ICEs.

Previously Bateman et al. (1997) showed the existence of a "part-whole bias" (PHB) in the monetary valuation of public goods (using contingent valuation methods). They showed that if the components are evaluated separately, the sum of those valuations tends to exceed the valuation placed on the whole. Our result has links to the PHB hypothesis in the sense that the sum of evaluations placed on the components in terms of WTP for each attribute (in the non-aggregation model) exceeds the value placed on the whole (META) in the aggregation model at different levels of information heterogeneity (Table 4.4). Looking at Table 4.4, the sum of WTP for each personalisation attribute outweighs the WTP for the whole attributes evaluated together (META) under different levels of information heterogeneity, suggesting that PHB can also exist in choice experiments.

Our results indicated that aggregating attributes of a good or service occur when more homogenous information is provided, impacting the WTP estimates. In this respect, CE practitioners should be aware of the possibility of aggregation when closely related attributes are presented and be more cautious in selecting attributes. The practitioners of CEs need to see the process of attribute development as very important as in the other elements of CE design. As an essential starting point in determining the nature of attributes, practitioners should follow the advice that attributes should be relevant to respondents and policymakers, and capable of being traded (Ryan, 1996). Attributes should not be too close to each other in terms of the information they provide; this avoids the possibility of being aggregated. Further, in one way or another, attributes selected should be differentiated from each other, as aggregation in CEs may bias results.

Our result may also be linked to the so-called 'support theory', a psychological model of a degree of belief, which assumes that the judged probability of an event generally increases when its description is unpacked into disjoint components. Rottenstreich and Tversky (1997) showed that when individuals are presented with an explicit disjunction (for instance, the probability that a particular student specialises in health economics, environmental economics, or agricultural economics), they may repack the various disciplines and evaluate the whole component 'economics' rather than the separate specialisations. The authors noted the presence of more explicit additivity for similar components than dissimilar components because similar parts are more easily repacked. However, whether individuals unpack or repack attributes in CEs should be further investigated.

Many factors (person-level and survey characteristics) affected AA behaviour. AA is more prevalent among participants who failed the monotonicity test. Monotonicity implies more of a desirable feature and less of an undesirable feature is preferred (Krucien et al., 2017). Respondents who failed the monotonicity test may not sufficiently consider all information provided and may have a limited attention to the attributes, hence are more likely to aggregate attributes. In health economics, Miguel et al. (2005), for example, tested for monotonicity and their finding implies that respondents who state they have great difficulty with a choice task are less likely to pass the monotonicity test, i.e. to choose the theoretically expected alternative. Not passing the monotonicity, therefore, may imply (difficulty of the tasks) or less attention to carefully consider each piece of information and hence more likely to adopt AA.

Task sequence (i.e., the position of the choice tasks within the questionnaire) also affected the aggregation threshold. Later positions of the choice tasks affect the aggregation threshold negatively. Respondents are more likely to adopt AA at the later positions of the choice tasks, potentially due to fatigue effect. Middle-positioned choice tasks affected the aggregation threshold positively, i.e., respondents are less likely to aggregate when choice tasks are paced in the middle-possibly-due to learning effect. This is consistent with Swait and Adamowicz (2001) who indicated an inverted bell-shaped effect of repeated task position on consistency, reflecting learning effects for an early position of the repeated choice task, and fatigue effects for later positions. In this regard, randomisation of the order of choice tasks across respondents may help minimise adoption of AA.

AA is also more likely to occur among those who spent a shorter time on completing the choice tasks. A shorter response time (RT) has a negative effect on the aggregation threshold, with participants using a quick and click strategy more likely to adopt AA. While it may be possible that respondents who answered relatively quickly processed all of the information in the choice tasks and made a utility-maximising choice, it is also likely that they utilised AA as a decision-making heuristic. Respondents that rush through the experiment (fast RT) may not sufficiently consider all information provided and hence are more likely to aggregate attributes. Holmes et al. (1998) find that respondents who took little time to answer the conjoint questions did not respond in ways that conform to underlying economic theory. We find evidence of a strong relationship between RT and the probability of adopting AA. For instance, for an information heterogeneity of equal to one (φ=1), the average probability of AA is 32%, but when choices are made quickly, this probability goes up to 82% and while it decreased to 12% when respondents took a longer time to complete choices, suggesting a very strong relationship between RT and AA. However, the relationship between RT and the tendency of adopting AA is not always straightforward as a longer RT may have other reasons than merely a high level of attention on the part of the respondent. It is possible that participants may engage in other activities on the computer.

Our study on AA could be extended in many ways. First, the estimation of the aggregation threshold (α) took place at the respondent level. However, one could also investigate changes in the threshold across choice tasks (i.e. $\alpha_n \rightarrow \alpha_{nt}$). Our approach could be adjusted to allow for potential dynamic changes in participants' decision-making. For example, participants may be less likely to adopt AA as a decision rule in the first few choice tasks, but as they go through the sequence of tasks a fatigue or boredom effect can make them more likely to adopt AA and then to lower the threshold value.

Second, the aggregation rule adopted gave the same importance to the four qualitative attributes in the aggregation process, consistent with the Dawes' rule (Dawes, 1979). The validity of this equal weighting hypothesis should be further explored. Future studies could make use of self-reported information about attributes importance to refine the weighting scheme.

Although this study underscores the importance of accommodating AA in choice modelling, there are some limitations. Firstly, whilst we focus on AA, we recognise that there are other decision heuristics not addressed in this paper. This would be particularly relevant if one intends to examine meaningful differences among various information processing strategies.

Secondly, there may be the risk of a potential confounding effect between AA as a decision rule and heterogeneity in preferences. We recognise the difficulties in separating heuristics and heterogeneity in preferences, that is, one cannot differentiate between the case where AA is the outcome of a simplifying decision-making heuristic and where it is instead a real indication of individual preferences.

Thirdly, we recognise that the cut-off points in determining AA as a form of binary classification of the information were arbitrarily selected. If the majority of qualitative attributes take high values (≥ 3), then the new dimension (or meta-attribute) would be "positive". Alternatively, if the majority of attributes take neutral values, then the new dimension would be "zero". Depending on the choice of the cut-off point[11], the binary classification approach is likely to underestimate either the actual number of attributes not aggregated or the number of attributes aggregated. Further research is required to determine the sensitivity of such results to the choice of the cut-off point or the form of aggregation (simple arithmetic versus binary classification of information for different cut-off points).

Finally, we used an opportunistic dataset to explore AA. Future research could incorporate appropriate supplementary questions to test the validity of our approach (for instance, by asking respondents whether they aggregate attributes or not).

4.7 Concluding remarks

Our results have important implications in terms of improving the design and choice modelling process for multi-attribute choices and guides practitioners in the analysis of CE data. We also highlight the need among CE practitioners and designers to broaden their scope of investigation beyond the standard behavioural assumptions. Future CEs should test for attributes processing strategies such as AA before using results for decision-making. Moreover, future research could use information processing-tracking techniques such as supplementary questions, think-aloud methods and an eye-tracker to understand the information processing strategies in CEs better.

[11]We also estimated the aggregation model with a cut-off point of 2 but failed to outperform the model with cut-off point of 3 (i.e., $AA - MNL_{Loglikelihood(cut-off=2)}$ =5861 versus $AA - MNL_{Loglikelihood(cut-off=3)}$ =5856; and $AA - MNL_{BIC(cut-off=2)}$ =11800.6 versus $AA - MNL_{BIC(cut-off=2)}$ = 11790.6).

Chapter 5

Conclusion

This thesis covers three independent studies in which health-related stated preferences are explored. All studies have in common that CE data is analysed while they address different research questions. The studies presented in chapters two and three exhibit some degree of similarity in that both are based on a stated preference survey concerned with preferences for time and risk attributes of kidney transplantation in Italy. The study in chapter four is based on a CE survey concerned with preferences for personalisation of chronic pain self-management programmes in the UK.

The objective of chapter two was to explore heterogeneity in patients' willingness to wait (WTW) for changes in time and risk attributes of kidney transplantation. Using the mixed logit model in WTW-space, the WTW parameters were directly estimated. We find heterogeneity in WTW for changes in the attributes of transplantation, and the variations in WTW correlate with two key observable characteristics of the patients, namely age and duration of dialysis. The model results indicated that younger patients are willing to wait longer for kidney transplantation with a better-expected outcome. Among patients on the waiting list, those who spent longer time on dialysis are willing to wait longer for a better quality kidney. The results highlighted the importance of accounting for observable characteristics of the patients in the design of kidney allocation protocols. Our results imply that patients' general welfare may be improved by embedding their preferences into the allocation algorithms.

The objective of the third chapter was to investigate whether there is a link between cognitive ability, choice consistency, and WTW for changes in time and risk attributes of kidney transplantation. Using the same dataset as in chapter two, the effect of patients' cognitive ability on the consistency of responses to the choice questions was analysed using heteroskedastic and generalised multinomial logit models. Patients with a higher cognitive ability responded more consistently to the choice questions. Whether the effect of cognitive ability, through choice consistency, is linked to WTW was analysed using the multinomial logit (MNL) model in which numeracy score (NS) interacts with the attributes. The interaction effect MNL model showed that patients with a higher cognitive ability are willing to wait less for kidney transplantation with a better-expected outcome. Here, the impact of cognitive ability on WTW is through choice consistency. That is, a higher cognitive ability resulted in a consistent response, and more consistent patients have a lower WTW.

The fourth chapter aimed to explore the presence of heuristics in the form of attributes aggregation (AA) in CEs. A crucial assumption underpinning a CE is that individuals consider all attributes and make a trade-off between them. However, attributes-based decision-making is cognitively demanding, often triggering the adoption of alternative decision rules. Using a non-linear utility model that allows attribute aggregation (AA) to depend on the information structure, we find that participants are more likely to aggregate information into a meta-attribute when the attributes provide similar information about the good or service. We show that accommodating AA as a choice simplifying rule in stated preferences outperforms the standard full attributes partitioning model. The probability of adopting AA is greater for homogenous information. Allowing for AA resulted in lower WTP estimates. Our results underline the importance of accounting for individuals' information processing rules when modelling multi-attribute choices.

5.1 Contribution of the thesis

Stated preference techniques have been extensively used within the health economics due to its solid theoretical foundation and as a means to measure preferences for various aspects of non-market goods or services or healthcare delivery. This PhD thesis has three main contributions:

1. **Investigation of heterogeneity in patients' willingness to wait for kidney transplantation**
   Patients' preferences for the time and risk characteristics of kidney transplantation has not been investigated before. The first paper shows that patients have heterogeneous preferences for the various attributes of an organ and that exploiting such heterogeneity would be very important from a policy perspective. Older patients showed a shorter WTW than younger ones. Patients with longer duration of dialysis have a higher WTW compared to patients at the early stage of dialysis. These results are important for policy-makers and should be taken into account in kid-

ney allocation protocols. However, before we conclude that marginal organs should be offered to the older candidates, or that patients with a longer time of dialysis exposure have a higher WTW and hence offer them a better quality kidney, this conclusion needs to be confirmed in reality. A follow-up study may be required to understand the dynamics of preferences and hence WTW estimates, but this is not within the scope of this paper.

2. **Demonstration of the effect of cognitive ability on choice consistency in a choice experiment**
   When eliciting preferences for some aspects of non-market goods or services, respondents are assumed to process each attribute separately without constraints. One of the problems with stated preference techniques is that respondents are presented with repeated choice tasks, and hence the possibility of making arbitrary choices is common -possibly due to limited or constrained information processing capacities. While measuring cognitive ability is likely to be relevant in many CEs, many studies have paid little attention to it. The second paper has shown the role of cognitive ability measures in identifying consistent responses. The paper indicated that patients with a higher cognitive ability are more likely to make a consistent choice decision and consistency leads to lower WTW for a better quality kidney. From a policy perspective, using inconsistent responses may lead to erroneous conclusions about WTW or WTP estimates. Therefore, if the goal of an experiment is to use such estimates for a cost-benefit analysis, accounting for the consistency of responses may be useful. However, further research in a different setting is required to confirm the result.

3. **Methodological contribution**
   In choice experiments (CEs) individuals are assumed to consider all the given attributes and make a trade-off between them. This allows estimation of marginal rates of substitution across individual attributes and willingness to pay (WTP) measures. However, attributes-based decision-making is cognitively challenging, and hence individuals may adopt alternative decision rules to decrease the cognitive difficulty of the choice task. Such simplifying decision rules have important consequences for the identification of the demand function as it implies a discontinuity in individuals' preferences and precludes computation of MRS. This thesis has shown one of the decision heuristics called attributes aggregation (AA), which has received very little attention in the CE literature. We find evidence of AA, and the probability of AA is higher for homogenous information. We have also shown the implications of AA on WTP estimates. We highlighted the importance of accounting for individuals' information processing strategies when modelling multi-attribute

choices. We focused on AA, but we recognise that there could be other decision heuristics not addressed in this paper. It may be interesting to compare AA with other information processing strategies. We leave the comparison as an avenue for further research.

Appendix A

Appendix

A.0.1 Kidney transplant survey (Original in Italian)

In what follows, the English translation of instructions and the questionnaire are presented.

I am part of a group of researchers from the University of Padua and the Ca' Foscari University of Venice carrying out a study that aims to assess whether it is possible to increase the well-being of patients who need a kidney transplant, naturally maintaining or by improving the clinical results of transplants. This research project, considered of strategic importance by the University of Padua, provides a survey on the characteristics and preferences of patients awaiting kidney transplantation. Your participation in this investigation is vital for scientific research. We will ask you about the preferences for alternative pairs of medical treatments, some demographic information, and your general state of health.

The results of this study will be published in specialised scientific journals and presented in scientific conferences. The information collected in this questionnaire will be linked to the information already held by the Regional Transplant Centre, but no publication or presentation will ever contain your name or any information that could identify you. All data collected will be archived and analysed in a strictly anonymous manner, pursuant to art. 7 and of the art. 13 of the Legislative Decree n. 196/03 in force since 1 January 2004 on the protection of individuals concerning the processing of personal data. Furthermore, the use of your data for commercial purposes is strictly prohibited. If you do not have any further questions or requests for clarification, we can start the interview.

Patients' preferences for the different transplant options

Instructions:

In this section sixteen alternative treatment pairs will be presented. You will be asked to express your preference between treatment A and treatment B by placing an X in the box below them. We remind you again that the answers will have no influence on how the future kidney transplant will be conducted. A transplant (treatment) is characterised by the following factors:

- Waiting time is the time one will have to wait in order to obtain the proposed transplant. The waiting time depends on the characteristics of the recipient and the frequency with which donors of a particular type are available.
- Graft survival is determined by the characteristics of the transplanted graft, the characteristics of the recipient, and the compatibility between donor and recipient.
- Infectious risk (standard or augmented) is the risk of contracting an infectious disease through the graft. If it is standard, the organ has undergone all the possible checks, even if complete safety cannot be guaranteed. If it is augmented, some of the controls have not been performed, or the donor had some risky behaviours in the days before his or her death, but an infection may still not result from clinical diagnostics (even if it is possible).
- Neoplastic risk (standard or augmented) is the risk of contracting a tumour through the transplanted organ. If it is standard, the donor was not affected by a tumour, almost surely, even if a minimum level of risk does exist (for example, if the donor was not aware of the problem and it did not emerge from checks). It is augmented if the donor had some kinds of neoplastic disease. Still, it is not high in terms of probability, because the due checks have been performed.

Below are proposed 16 pairs of treatments (transplants) described by different attributes. Please, indicate the preferred one for each pair, by crossing (X) in the square below it.

1. Which of the two treatments would you prefer? Put an X below the chosen treatment

	Treatment A	**Treatment B**
Waiting Time	6 Months	6 Months
Expected Graft Survival	20 Years	15 Years
Infectious Risk	Standard	Standard
Neoplastic Risk	Augmented	Standard
Your Choice ?	☐	☐

2. Which of the two treatments would you prefer? Put an X below the chosen treatment

	Treatment A	**Treatment B**
Waiting Time	12 Months	36 Months
Expected Graft Survival	15 Years	20 Years
Infectious Risk	Standard	Augmented
Neoplastic Risk	Standard	Augmented
Your Choice ?	☐	☐

3. Which of the two treatments would you prefer? Put an X below the chosen treatment

	Treatment A	**Treatment B**
Waiting Time	60 Months	6 Months
Expected Graft Survival	20 Years	15 Years
Infectious Risk	Standard	Augmented
Neoplastic Risk	Augmented	Standard
Your Choice ?	☐	☐

4. Which of the two treatments would you prefer? Put an X below the chosen treatment

	Treatment A	**Treatment B**
Waiting Time	6 Months	12 Months
Expected Graft Survival	10 Years	10 Years
Infectious Risk	Augmented	Standard
Neoplastic Risk	Augmented	Augmented
Your Choice ?	☐	☐

5. Which of the two treatments would you prefer? Put an X below the chosen treatment

	Treatment A	**Treatment B**
Waiting Time	36 Months	60 Months
Expected Graft Survival	10 Years	10 Years
Infectious Risk	Augmented	Standard
Neoplastic Risk	Standard	Standard
Your Choice ?	☐	☐

6. Which of the two treatments would you prefer? Put an X below the chosen treatment

	Treatment A	**Treatment B**
Waiting Time	60 Months	36 Months
Expected Graft Survival	15 Years	10 Years
Infectious Risk	Augmented	Augmented
Neoplastic Risk	Augmented	Standard
Your Choice ?	☐	☐

7. Which of the two treatments would you prefer? Put an X below the chosen treatment

	Treatment A	**Treatment B**
Waiting Time	60 Months	60 Months
Expected Graft Survival	20 Years	20 Years
Infectious Risk	Augmented	Standard
Neoplastic Risk	Standard	Augmented
Your Choice ?	☐	☐

8. Which of the two treatments would you prefer? Put an X below the chosen treatment

	Treatment A	**Treatment B**
Waiting Time	36 Months	6 Months
Expected Graft Survival	15 Years	10 Years
Infectious Risk	Standard	Augmented
Neoplastic Risk	Augmented	Augmented
Your Choice ?	☐	☐

9. Which of the two treatments would you prefer? Put an X below the chosen treatment

	Treatment A	**Treatment B**
Waiting Time	6 Months	12 Months
Expected Graft Survival	15 Years	20 Years
Infectious Risk	Standard	Augmented
Neoplastic Risk	Standard	Standard
Your Choice ?	☐	☐

10.Which of the two treatments would you prefer? Put an X below the chosen treatment

	Treatment A	**Treatment B**
Waiting Time	12 Months	60 Months
Expected Graft Survival	10 Years	15 Years
Infectious Risk	Standard	Augmented
Neoplastic Risk	Augmented	Augmented
Your Choice ?	☐	☐

11.Which of the two treatments would you prefer? Put an X below the chosen treatment

	Treatment A	**Treatment B**
Waiting Time	12 Months	36 Months
Expected Graft Survival	20 Years	20 Years
Infectious Risk	Augmented	Standard
Neoplastic Risk	Standard	Standard
Your Choice ?	☐	☐

12.Which of the two treatments would you prefer? Put an X below the chosen treatment

	Treatment A	**Treatment B**
Waiting Time	6 Months	12 Months
Expected Graft Survival	15 Years	15 Years
Infectious Risk	Augmented	Standard
Neoplastic Risk	Standard	Standard
Your Choice ?	☐	☐

13.Which of the two treatments would you prefer? Put an X below the chosen treatment

	Treatment A	**Treatment B**
Waiting Time	60 Months	12 Months
Expected Graft Survival	10 Years	15 Years
Infectious Risk	Standard	Augmented
Neoplastic Risk	Standard	Augmented
Your Choice ?	☐	☐

14.Which of the two treatments would you prefer? Put an X below the chosen treatment

	Treatment A	**Treatment B**
Waiting Time	36 Months	60 Months
Expected Graft Survival	20 Years	20 Years
Infectious Risk	Augmented	Augmented
Neoplastic Risk	Augmented	Standard
Your Choice ?	☐	☐

15.Which of the two treatments would you prefer? Put an X below the chosen treatment

	Treatment A	**Treatment B**
Waiting Time	36 Months	6 Months
Expected Graft Survival	20 Years	20 Years
Infectious Risk	Standard	Standard
Neoplastic Risk	Standard	Augmented
Your Choice ?	☐	☐

16.Which of the two treatments would you prefer? Put an X below the chosen treatment

	Treatment A	**Treatment B**
Waiting Time	12 Months	36 Months
Expected Graft Survival	15 Years	15 Years
Infectious Risk	Augmented	Standard
Neoplastic Risk	Augmented	Augmented
Your Choice ?	☐	☐

We thank you for your precious time and collaboration. Next are a few questions about the logical abilities of patients about different combinations of choices.

SHARE Numeracy Questions

Now I would like to ask you some questions that are needed to evaluate how people use numbers in everyday life.

1. The probability of contracting an illness is 10 percent, how many people out of one thousand would be expected to get the disease?
2. In a sale, a shop is selling all items at half price. Before the sale the sofa costs 300 Euros. How much will it cost in the sale?
3. A second hand car dealer is selling a car for 6,000 Euro. This is two-thirds of what it costs new. How much did the car cost new?

Personal information:

1. **Education:**

   ☐ Elementary ☐ Lower middle ☐ Higher middle ☐ Degree
2. **Family composition (not just the people living with you)**

   ☐ Mother ☐ Father ☐ Brothers/sisters ☐ Male-No.———- ☐ Female-No.——— ☐ Wife ☐ Husband ☐ Cohabiting ☐ Children ☐ Male-No.———- ☐ Female-No.———-☐ Other
3. **What is your current profession?**

   ☐ Manager ☐ Freelancer ☐ Worker ☐ Housewife ☐ Retired ☐ Student ☐ Other——
4. **Do you currently have a disability pension?**

   ☐ Yes ☐ No

Medical information:

1. First year diagnosis/age of onset of the pathology———-
2. Dialysis start date: month/year———-
3. Dialysis type

   ☐ Haemodialysis ☐ Peritoneal dialysis
4. Presence of diabetes mellitus
   ☐ yes ☐ no
5. Date listed for renal transplantation: ——/——/ ——

Dialysis:

In your opinion, how true or false are the following statements?

		Absolutely True	**True**	**I don't know**	**False**	**Absolutely False**
1	Dialysis affects my life too much	1	2	3	4	5
2	Dialysis makes me lose too much time	1	2	3	4	5
3	I find it frustrating to live with dialysis	1	2	3	4	5
4	I feel dialysis a burden to my family	1	2	3	4	5

General health status:

☐ Excellent ☐ Very good ☐ Good ☐ Passable ☐ Poor

Bibliography

Agarwal, S. and Mazumder, B. (2013). Cognitive abilities and household financial decision making. *American Economic Journal: Applied Economics*, 5(1):193–207.

Aizaki, H. and Nishimura, K. (2008). Design and analysis of choice experiments using r: a brief introduction. *Agricultural information research*, 17(2):86–94.

Alemu, M. H., Mørkbak, M. R., Olsen, S. B., and Jensen, C. L. (2013). Attending to the reasons for attribute non-attendance in choice experiments. *Environmental and resource economics*, 54(3):333–359.

Banks, J. and Oldfield, Z. (2007). Understanding pensions: Cognitive function, numerical ability and retirement saving. *Fiscal studies*, 28(2):143–170.

Barbara, F., Harry, T., and Peter, Z. (2018). End-of-life healthcare expenditure: testing economic explanations using a discrete choice experiment. *Journal of health economics*.

Bateman, I., Munro, A., Rhodes, B., Starmer, C., and Sugden, R. (1997). Does part–whole bias exist? an experimental investigation. *The Economic Journal*, 107(441):322–332.

Battiston, G., Nicoló, A., Pace, N., and Pasini, G. (2016). Incorporating preference axioms in discrete choice experiment design. mimeo.

Bech, M., Gyrd-Hansen, D., Kjær, T., Lauridsen, J., and Sørensen, J. (2007). Graded pairs comparison-does strength of preference matter? analysis of preferences for specialised nurse home visits for pain management. *Health economics*, 16(5):513–529.

Bech, M., Kjaer, T., and Lauridsen, J. (2011). Does the number of choice sets matter? results from a web survey applying a discrete choice experiment. *Health economics*, 20(3):273–286.

Benjamin, D. J., Brown, S. A., and Shapiro, J. M. (2013). Who is ?behavioral?? cognitive ability and anomalous preferences. *Journal of the European Economic Association*, 11(6):1231–1255.

Bhat, C. R. (2000). Incorporating observed and unobserved heterogeneity in urban work travel mode choice modeling. *Transportation science*, 34(2):228–238.

Bliemer, M. C. and Rose, J. M. (2005a). Efficiency and sample size requirements for stated choice studies. *Institute of Transport Studies and Logistics Working Paper*, (ITLS-WP-05-08).

Bliemer, M. C. and Rose, J. M. (2005b). Efficient designs for alternative specific choice experiments.

Brañas-Garza, P., Garcia-Munoz, T., and González, R. H. (2012). Cognitive effort in the beauty contest game. *Journal of Economic Behavior & Organization*, 83(2):254–260.

Brown, P., Panattoni, L., Cameron, L., Knox, S., Ashton, T., Tenbensel, T., and Windsor, J. (2015). Hospital sector choice and support for public hospital care in new zealand: Results from a labeled discrete choice survey. *Journal of health economics*, 43:118–127.

Burks, S. V., Carpenter, J. P., Goette, L., and Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences*, 106(19):7745–7750.

Burton, C. D., Entwistle, V. A., Elliott, A. M., Krucien, N., Porteous, T., and Ryan, M. (2017). The value of different aspects of person-centred care: a series of discrete choice experiments in people with long-term conditions. *BMJ open*, 7(4):e015689.

Campbell, D., Hensher, D. A., and Scarpa, R. (2011). Non-attendance to attributes in environmental choice analysis: a latent class specification. *Journal of environmental planning and management*, 54(8):1061–1076.

Cappelletti, D., Güth, W., and Ploner, M. (2011). Being of two minds: Ultimatum offers under cognitive constraints. *Journal of Economic Psychology*, 32(6):940–950.

Christelis, D., Jappelli, T., and Padula, M. (2010). Cognitive abilities and portfolio choice. *European Economic Review*, 54(1):18–38.

Clark, M. D., Gumber, A., Leech, D., Moro, D., Szczepura, A. K., West, N., and Higgins, R. M. (2009). Prioritizing patients for renal transplantation?: Analysis of patient preferences for kidney allocation according to ethnicity and gender. *Diversity and Equality in Health and Care*, 6(3):181–191.

Clark, M. D., Leech, D., Gumber, A., Moro, D., Szczepura, A., West, N., and Higgins, R. (2012). Who should be prioritized for renal transplantation?: Analysis of key stakeholder preferences using discrete choice experiments. *BMC nephrology*, 13(1):1.

Clark, M. D., Szczepura, A., Gumber, A., Howard, K., Moro, D., and Morton, R. L. (2018). Measuring trade-offs in nephrology: a systematic review of discrete choice experiments and conjoint analysis studies. *Nephrology Dialysis Transplantation*, 33(2):348–355.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., and Garcia-Retamero, R. (2012). Measuring risk literacy: The berlin numeracy test. *Judgment and Decision Making*.

Cokely, E. T. and Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation.

Courtney, A. E. and Maxwell, A. P. (2008). The challenge of doing what is right in renal transplantation: balancing equity and utility. *Nephron Clinical Practice*, 111(1):c62–c68.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American psychologist*, 34(7):571.

de Bekker-Grob, E. W., Ryan, M., and Gerard, K. (2012). Discrete choice experiments in health economics: a review of the literature. *Health economics*, 21(2):145–172.

Deck, C. and Jahedi, S. (2015). The effect of cognitive load on economic decision making: A survey and new experiments. *European Economic Review*, 78:97–119.

DeShazo, J. and Fermo, G. (2002). Designing choice sets for stated preference methods: the effects of complexity on choice consistency. *Journal of Environmental Economics and management*, 44(1):123–143.

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, 100(3):1238–60.

Erdem, S., Campbell, D., and Hole, A. R. (2015). Accounting for attribute-level non-attendance in a health choice experiment: does it matter? *Health economics*, 24(7):773–789.

Erdem, S., Campbell, D., and Thompson, C. (2014). Elimination and selection by aspects in health choice experiments: prioritising health service innovations. *Journal of health economics*, 38:10–22.

Fiebig, D. G., Keane, M. P., Louviere, J., and Wasi, N. (2010). The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Marketing Science*, 29(3):393–421.

Foote, C., Morton, R. L., Jardine, M., Gallagher, M., Brown, M., Howard, K., and Cass, A. (2014). Considerations of nephrologists when suggesting dialysis in elderly patients with renal failure (consider): a discrete choice experiment. *Nephrology Dialysis Transplantation*, 29(12):2302–2309.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, 19(4):25–42.

Genie, M., Nicoló, A., and Pasini, G. (2018). Heterogeneity in patients' willingness to wait for attributes of kidney transplantation: Does it matter? mimeo.

Gilbride, T. J. and Allenby, G. M. (2004). A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science*, 23(3):391–406.

Graffeo, M., Polonio, L., and Bonini, N. (2015). Individual differences in competent consumer choice: the role of cognitive reflection and numeracy skills. *Frontiers in psychology*, 6:844.

Greene, W. H. and Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological*, 37(8):681–698.

Greene, W. H., Hensher, D. A., and Rose, J. (2006). Accounting for heterogeneity in the variance of unobserved effects in mixed logit models. *Transportation Research Part B: Methodological*, 40(1):75–92.

Gu, Y., Hole, A. R., Knox, S., et al. (2013). Fitting the generalized multinomial logit model in stata. *Stata J*, 13(2):382–397.

Hagemi, A., Plumpton, C., and Hughes, D. A. (2017). Renal transplant patients? preference for the supply and delivery of immunosuppressants in wales: a discrete choice experiment. *BMC nephrology*, 18(1):305.

Halpern, S. D., Berns, J. S., and Israni, A. K. (2004). Willingness of patients to switch from conventional to daily hemodialysis: looking before we leap. *The American journal of medicine*, 116(9):606–612.

Heidenreich, S., Watson, V., Ryan, M., and Phimister, E. (2018). Decision heuristic or preference? attribute non-attendance in discrete choice problems. *Health economics*, 27(1):157–171.

Hensher, D. A. (2006). How do respondents process stated choice experiments? attribute consideration under varying information load. *Journal of applied econometrics*, 21(6):861–878.

Hess, S. and Rose, J. M. (2012). Can scale and coefficient heterogeneity be separated in random coefficients models? *Transportation*, 39(6):1225–1239.

Hole, A. (2009). Clogithet: Stata module to estimate heteroscedastic conditional logit model.

Hole, A. R. (2008). Modelling heterogeneity in patients preferences for the attributes of a general practitioner appointment. *Journal of health economics*, 27(4):1078–1094.

Hole, A. R. (2011). Attribute non-attendance in patients? choice of general practitioner appointment. In *second international choice modelling conference, Oulton Hall, Leeds*. Citeseer.

Hole, A. R. et al. (2006). Small-sample properties of tests for heteroscedasticity in the conditional logit model. *Economics Bulletin*, 3(18):1–14.

Hole, A. R. and Kolstad, J. R. (2012). Mixed logit estimation of willingness to pay distributions: a comparison of models in preference and wtp space using data from a health-related choice experiment. *Empirical Economics*, 42(2):445–469.

Hole, A. R., Kolstad, J. R., and Gyrd-Hansen, D. (2013). Inferred vs. stated attribute non-attendance in choice experiments: A study of doctors? prescription behaviour. *Journal of Economic Behavior & Organization*, 96:21–31.

Holmes, T., Alger, K., Zinkhan, C., and Mercer, D. E. (1998). The effect of response time on conjoint analysis estimates of rainforest protection values. *Journal of Forestry Economics 4: 1 1998.*

Howard, K., Jan, S., Rose, J. M., Wong, G., Craig, J. C., Irving, M., Tong, A., Chadban, S., Allen, R. D., and Cass, A. (2016). Preferences for policy options for deceased organ donation for transplantation: a discrete choice experiment. *Transplantation*, 100(5):1136–1148.

Howard, K., Jan, S., Rose, J. M., Wong, G., Irving, M., Tong, A., Craig, J. C., Chadban, S., Allen, R. D., and Cass, A. (2015). Community preferences for the allocation of donor organs for transplantation: a discrete choice study. *Transplantation*, 99(3):560–567.

Kamran, S., Conti, F., Pomey, M.-P., Baron, G., Calmus, Y., and Vidal-Trecan, G. (2017). Patients? preferences in transplantation from marginal donors: results of a discrete choice experiment. *Transplant International*, 30(6):589–602.

Kessels, R., Goos, P., and Vandebroek, M. (2006). A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research*, 43(3):409–419.

Kjær, T., Bech, M., Kronborg, C., and Mørkbak, M. R. (2013). Public preferences for establishing nephrology facilities in greenland: estimating willingness-to-pay using a discrete choice experiment. *The European Journal of Health Economics*, 14(5):739–748.

Kolstad, J. R. (2011). How to make rural jobs more attractive to health workers. findings from a discrete choice experiment in tanzania. *Health economics*, 20(2):196–211.

Krucien, N., Watson, V., and Ryan, M. (2017). Is best–worst scaling suitable for health state valuation? a comparison with discrete choice experiments. *Health economics*, 26(12):e1–e16.

Lagarde, M. (2013). Investigating attribute non-attendance and its consequences in choice experiments with latent class models. *Health economics*, 22(5):554–567.

Lancaster, K. J. (1966). A new approach to consumer theory. *The journal of political economy*, pages 132–157.

Lancsar, E. and Louviere, J. (2008). Conducting discrete choice experiments to inform healthcare decision making. *Pharmacoeconomics*, 26(8):661–677.

LaRiviere, J., Czajkowski, M., Hanley, N., Aanesen, M., Falk-Petersen, J., and Tinch, D. (2014). The value of familiarity: effects of knowledge and objective signals on willingness to pay for a public good. *Journal of Environmental Economics and Management*, 68(2):376–389.

Layton, D. and Hensher, D. A. (2010). Aggregation of common-metric attributes in preference revelation and implications for willingness to pay. *Transportation Research Part D: Transport and Environment*, 15(7):394–404.

Lipkus, I. M., Samsa, G., and Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical decision making*, 21(1):37–44.

Louviere, J. J., Hensher, D. A., and Swait, J. D. (2000). *Stated choice methods: analysis and applications*. Cambridge university press.

Louviere, J. J. and Timmermans, H. J. (1990). Using hierarchical information integration to model consumer responses to possible planning actions: recreation destination choice illustration. *Environment and Planning A*, 22(3):291–308.

Manski, C. F. (1977). The structure of random utility models. *Theory and decision*, 8(3):229–254.

Marshall, D., Deal, K., Conner-Spady, B., Bohm, E., Hawker, G., Loucks, L., MacDonald, K., and Noseworthy, T. (2018). How do patients trade-off surgeon choice and waiting times for total joint replacement: a discrete choice experiment. *Osteoarthritis and cartilage*, 26(4):522–530.

Matas, A., Smith, J., Skeans, M., Thompson, B., Gustafson, S., Stewart, D., Cherikh, W., Wainright, J., Boyle, G., Snyder, J., et al. (2015). Optn/srtr 2013 annual data report: kidney. *American journal of transplantation*, 15(S2):1–34.

Mazzonna, F. and Peracchi, F. (2012). Ageing, cognitive abilities and retirement. *European Economic Review*, 56(4):691–710.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. pages 105–142.

McFadden, D. and Train, K. (2000). Mixed mnl models for discrete response. *Journal of applied Econometrics*, pages 447–470.

Merion, R. M., Ashby, V. B., Wolfe, R. A., Distant, D. A., Hulbert-Shearon, T. E., Metzger, R. A., Ojo, A. O., and Port, F. K. (2005). Deceased-donor characteristics and the survival benefit of kidney transplantation. *Jama*, 294(21):2726–2733.

Metzger, R. A., Delmonico, F. L., Feng, S., Port, F. K., Wynn, J. J., and Merion, R. M. (2003). Expanded criteria donors for kidney transplantation. *American Journal of Transplantation*, 3(s4):114–125.

Miguel, F. S., Ryan, M., and Amaya-Amaya, M. (2005). ?irrational?stated preferences: a quantitative and qualitative investigation. *Health economics*, 14(3):307–322.

Milte, R., Ratcliffe, J., Chen, G., Lancsar, E., Miller, M., and Crotty, M. (2014). Cognitive overload? an exploration of the potential impact of cognitive functioning in discrete choice experiments with older people in health care. *Value in Health*, 17(5):655–659.

Molin, E. J. and Timmermans, H. J. (2009). Hierarchical information integration experiments and integrated choice experiments. *Transport reviews*, 29(5):635–655.

Morton, R. L., Snelling, P., Webster, A. C., Rose, J., Masterson, R., Johnson, D. W., and Howard, K. (2012). Dialysis modality preference of patients with ckd and family caregivers: a discrete-choice study. *American Journal of Kidney Diseases*, 60(1):102–111.

Nieboer, A. P., Koolman, X., and Stolk, E. A. (2010). Preferences for long-term care services:

Willingness to pay estimates derived from a discrete choice experiment. *Social science & medicine*, 70(9):1317–1325.

Ojo, A. O., Hanson, J. A., Meier-Kriesche, H.-U., Okechukwu, C. N., Wolfe, R. A., Leichtman, A. B., Agodoa, L. Y., Kaplan, B., and Port, F. K. (2001). Survival in recipients of marginal cadaveric donor kidneys compared with other recipients and wait-listed transplant candidates. *Journal of the American Society of Nephrology*, 12(3):589–597.

Oppewal, H., Louviere, J. J., and Timmermans, H. J. (1994). Modeling hierarchical conjoint processes with integrated choice experiments. *Journal of Marketing Research*, pages 92–105.

Özdemir, S., Johnson, F. R., and Hauber, A. B. (2009). Hypothetical bias, cheap talk, and stated willingness to pay for health care. *Journal of health economics*, 28(4):894–901.

Pak, T.-Y. and Babiarz, P. (2018). Does cognitive aging affect portfolio choice? *Journal of Economic Psychology*, 66:1–12.

Peters, E. and Levin, I. P. (2008). Dissecting the risky-choice framing effect: Numeracy as an individual-difference factor in weighting risky and riskless options.

Reese, P. P., Tehrani, T., Lim, M. A., Asch, D. A., Blumberg, E. A., Simon, M. K., Bloom, R. D., and Halpern, S. D. (2010). Determinants of the decision to accept a kidney from a donor at increased risk for blood-borne viral infection. *Clinical Journal of the American Society of Nephrology*, pages CJN–08251109.

Roth, A. E., Sönmez, T., and Ünver, M. U. (2004). Kidney exchange. *The Quarterly Journal of Economics*, 119(2):457–488.

Rottenstreich, Y. and Tversky, A. (1997). Unpacking, repacking, and anchoring: advances in support theory. *Psychological review*, 104(2):406.

Rousseau, S. and Rousseau, R. (2012). Interactions between journal attributes and authors' willingness to wait for editorial decisions. *Journal of the American Society for Information Science and Technology*, 63(6):1213–1225.

Ryan, M. and Gerard, K. (2003). Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Applied health economics and health policy*, 2(1):55–64.

Ryan, M., Gerard, K., and Amaya-Amaya, M. (2007). *Using discrete choice experiments to value health and health care*, volume 11. Springer Science & Business Media.

Ryan, M., Gerard, K., and Amaya-Amaya, M. (2008). Discrete choice experiments in a nutshell. In *Using discrete choice experiments to value health and health care*, pages 13–46. Springer.

Salvioli, M., Lucchetti, R., and Torelli, R. (2016). Simulating the impact of crossover kidney transplantation on the nord italia transplant program. *Games*, 7(4):30.

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, 5(17):61–71.

Scarpa, R., Thiene, M., and Train, K. (2006). Utility in wtp space: a tool to address confounding random scale effects in destination choice to the alps.

Scarpa, R., Thiene, M., and Train, K. (2008). Utility in willingness to pay space: a tool to address confounding random scale effects in destination choice to the alps. *American Journal of Agricultural Economics*, 90(4):994–1010.

Schold, J. D. and Meier-Kriesche, H.-U. (2006). Which renal transplant candidates should accept marginal kidneys in exchange for a shorter waiting time on dialysis? *Clinical Journal of the American Society of Nephrology*, 1(3):532–538.

Sefora, P. E., Silvio, S., Nicola, D. F., Giuseppe, R., Iris, F., Luigino, B., Maria, G., Eliana, G., Donato, D., Enrico, M., et al. (2013). Optimizing utilization of kidneys from deceased donors over 60 years: Five-year outcomes after implementation of a combined clinical and histological allocation algorithm. *Transplant International*, 26(8):833–841.

Shugan, S. M. (1980). The cost of thinking. *Journal of consumer Research*, 7(2):99–111.

Swait, J. and Adamowicz, W. (2001). Choice environment, market complexity, and consumer behavior: a theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organizational Behavior and Human Decision Processes*, 86(2):141–167.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological review*, 34(4):273.

Train, K. and Weeks, M. (2005). Discrete choice models in preference space and willingness-to-pay space. *Applications of simulation methods in environmental and resource economics*, pages 1–16.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological review*, 79(4):281.

van Helvoort-Postulart, D., Dellaert, B. G., Van Der Weijden, T., Von Meyenfeldt, M. F., and Dirksen, C. D. (2009). Discrete choice experiments for complex health-care decisions: does hierarchical information integration offer a solution? *Health economics*, 18(8):903–920.

Vatn, A. and Bromley, D. W. (1994). Choices without prices without apologies. *Journal of environmental economics and management*, 26(2):129–148.

Veldwijk, J., Lambooij, M. S., de Bekker-Grob, E. W., Smit, H. A., and De Wit, G. A. (2014). The effect of including an opt-out option in discrete choice experiments. *PloS one*, 9(11):e111805.

Watson, V., Ryan, M., Brown, C., Barnett, G., Ellis, B., and Emberton, M. (2004). Eliciting preferences for drug treatment of lower urinary tract symptoms associated with benign prostatic hyperplasia. *The Journal of urology*, 172(6):2321–2325.

Wheeler, B. (2006). Algdesign.

Whiting, J., Zavala, E., Alexander, J., and First, M. (1999). The cost-effectiveness of transplantation with expanded donor kidneys. In *Transplantation proceedings*, volume 31, pages 1320–1321.

Wright, S. J., Vass, C. M., Sim, G., Burton, M., Fiebig, D. G., and Payne, K. (2018). Accounting for scale heterogeneity in healthcare-related discrete choice experiments when comparing stated preferences: A systematic review. *The Patient-Patient-Centered Outcomes Research*, pages 1–14.