Eliciting preferences using Discrete Choice Experiments in Healthcare: Willingness to Pay, Stakeholder preferences, and Altruistic preferences.

By

Michael David Clark B.Sc Hons (Economics); M.Sc (Health Economics); M.Sc (Quantitative Development Economics).

Submission for the degree of Doctor of Philosophy (PhD).

Registered within the Department of Economics, (However the research was undertaken whilst working for Warwick Business School, and more recently Warwick Medical School).

Submitted December 2013.

Table of contents

Chapter 6 (Prioritizing patients for renal transplantation? Catering for diversity by analyzing patient preferences for kidney allocation according to ethnicity and gender). P254

Appendices.

List of figures and tables.

Chapter 1.

Chapter 2.

Chapter 3.

Chapter 4.

Chapter 5.

Chapter 6.

Chapter 7.

Chapter 8

Acknowledgements.

I want to acknowledge and thank a lot of different people for their support. I am particularly grateful to Professor Dennis Leech for doing an outstanding job of supervising this PhD, despite the fact that he is not a Health Economist. His particular expertise in Econometrics has been invaluable. Professor Ala Szczepura has been a constant source of encouragement and support, and has acted as a sort of informal second supervisor with respect to Chapters 2-4 of the thesis, she has also proof read some chapters of the thesis. Dr Emma Mc Intosh initially introduced me to DCE techniques, and I am very grateful to her for that. Dr Domenico Moro has been a constant source of help in terms of familiarizing me with applications packages such as STATA and also teaching me how to conduct some of the final econometric analysis. The people who have collaborated with me in this research (see section 'Declaration of research collaboration') also ought to be acknowledged and thanked. I am also conscious of the fact that people who have supported me whilst I took earlier degrees including Dr David Burningham (Brunel University), and Dr Karl Claxton (University of York), and Dr Alan Roe (University of Warwick), played a major part in ensuring I was in a position to take this PhD in the first instance so thanks are also due to them. Finally of course support from my family has been invaluable.

Declaration.

I declare that this is my own work, although some of it is based upon permitted collaborative research. For details see 'Declaration of Research collaboration.' This material has not been submitted as part of any other degree at another University. It should be noted that I reported preliminary results from the analysis in chapter 2 for my M.Sc dissertation at the University of Warwick. The pilot stage of the analysis reported in chapter 2 was therefore undertaken before I embarked on my PhD. However, the final data analysis reported here (on a larger sample of respondents), the literature review, and writing up of the published paper, and Chapter 2 have all been undertaken after my M.Sc, and during my period of PhD registration.

Abbreviations.

DCE: Discrete Choice Experiment.

DCEs: Discrete Choice Experiments.

DVT: Deep Vein Thombosis.

Eq-5d: Euroqol 5 dimensional Quality of Life Scale.

HLA: Human Leucocyte Antigen.

LCM: Latent Class Model

NHS: National Health Service.

VAS: Visual Analogue Scale.

WTP: Willingness to Pay.

Declaration of research collaboration.

Chapter 1 of the thesis was written by me. However, my supervisor Professor Dennis Leech kindly looked through it, and suggested some edits to improve the continuity of the text. I took these suggestions on board and made appropriate edits. I met with Professor Stavros Petrou to discuss the literature review in section D of the chapter on a few occasions. Dr Domenico Moro also looked at it, but could find no errors to draw to my attention. Moreover, Professor Ala Szczepura proof read the final chapter, and suggested some stylistic edits.

Much of the analysis in Chapter 2 of the thesis has been published as paper by Health Policy. However, the chapter also contains analysis relating to uptake rates and compensating variation, which is not included in the paper. I am the first author of the associated paper reflecting the fact that I conducted the final data analysis, conducted literature searches, reviewed the literature, and wrote the paper. Professor Ala Szczepura looked at my draft and suggested some stylistic changes. Dr Domenico Moro is also an author reflecting that fact that the data analysis for the first submission to the journal was conducted by him (it involved the use of bootstrapped confidence intervals for Willingness to Pay which he had estimated). However a peer reviewer suggested that we provide a simple explanation of the econometric methods used. I therefore re-analyzed the data myself using STATA version 9.2, and the Delta method for the confidence intervals (as this could be explained more easily). As per the acknowledgements in the published paper, because this was my first DCE I enlisted the support of Dr Emma Mc Intosh, who taught me how to use SPEED for the purposes of the DCE design. She also showed me how to analyze the early preliminary datasets using LIMDEP (the pilot analysis on a smaller dataset also formed the basis of my M.Sc dissertation for my second M.Sc at Warwick University). Emma has had no involvement in the final data analysis or with the final paper or chapter 2. I received assistance from medical

professionals in Leicester relating to the selection of appropriate attributes and levels during piloting, and with the distribution and collection of questionnaires. The authorship for the paper was listed as: Mr Michael D. Clark, Dr Domenico Moro, and Professor Ala K. Szczepura.

Chapter 3 of the thesis relates to the application of DCEs to Gynecology provision. The initial process of selecting attributes for the pilot questionnaire arose as a result of discussions between me and clinicians working on the project. I then designed the pilot questionnaires using SPEED, and designed a ranking exercise whereby we asked respondents to rank potential attributes in order of priority. This was so that we could consider other potential attributes for inclusion in the final questionnaire design. The actual process of piloting was conducted by Dr Sophia Julian at Leicester Royal Infirmary. She fed back to me on the results of the ranking exercise and obtained responses to the pilot questionnaires. I then analyzed this data using STATA. There was then some considerable debate between various members of the project team about the selection of attributes and levels, but I worked to ensure that attributes and levels selected seemed sensible. I then produced a revised version of the questionnaire using SPEED which was distributed by Dr Julian to patients in Leicester. I conducted the final data analysis and wrote Chapter 3. The idea of including a question to explore whether respondents took a monetary attribute into account (in chapters 3 and 4) was mine. I received guidance particularly in relation to the use of econometric notation from Professor Leech my supervisor. Moreover Dr Marwan Habiba, Dr Sophia Julian, and Professor Ala Szczepura and Professor Dennis Leech, made comments and suggested minor changes to the chapter. Dr Moro introduced me to the use of STATA 9.2 and the Delta method using a different dataset. He has done a lot of data analysis on sub-group data emanating from this project which is not reported within this paper.

Chapter 4 of the thesis reports on results from a DCE which was essentially the same as the DCE used in chapter 3 (although the descriptor for the monetary attribute was varied between different versions of the questionnaire). The piloting and development process was the same one as for the analysis reported in chapter 3. The idea of conducting an analysis to establish whether using different descriptors for the monetary attribute might influence estimated WTP was my own. Once again I received guidance particularly in relation to the use of econometric notation from Professor Leech. Also Dr Marwan Habiba, Dr Sophia Julian, and Professor Ala Szczepura and Professor Dennis Leech suggested minor changes to drafts of the chapter. I undertook the data analysis for this paper myself unaided.

For the renal transplant project (Chapters 5, 6, 7, and 8 of the thesis) I worked with physicians to work out what the attributes might be for the DCE. I then generated a pilot DCE questionnaire using SPEED. Dr Julie Ratcliffe provided input on a consultancy basis with a view to me learning how to use SPSS to generate a final design template for the final DCE. However, in the end because she had technical problems with SPSS she sourced a design template from leaders in the field instead (Street et al. 2005). Julie provided some advice on coding for analysis of the pilot data, and confirmed that the final selection of attributes seemed reasonable, but has had no further involvement in the project since the end of the piloting stage.

I conducted about 50 / 60 of the interviews during the piloting stage. Dr Anil Gumber conducted the remainder because he is bilingual, and could therefore interview ethnic minorities. The piloting stage involved respondents ranking attributes, having the opportunity to suggest others, and then filling in a questionnaire. After piloting I analyzed the data using STATA. I then organized the printing and mailing out of flyers for enclosure in the publication Kidney Life (20,000 flyers). When responses came back (with clerical support) I ensured that

we sent out questionnaires to the patient, carer, donor, and healthcare worker respondents. In addition to this I organized a separate mailing to live donors and relatives of deceased donors via the British Organ Donor Society (BODY). The mailing to healthcare workers was organized by Nicholas West (transplant co-coordinator) and Dr Robert Higgins from the University Hospital in Coventry. Dr Anil Gumber arranged for questionnaires to be translated into 5 other languages for non-English speaking respondents. To ensure we had a large enough sample of respondents from ethnic minority groups, he also visited hospitals to obtain responses from ethnic minorities. Dr Domenico Moro was involved in early analysis of the data using STATA v. 7, but all the analysis in chapters 5, 6, and 8 was conducted by me using STATA 9.2 (which allows for the calculation of confidence intervals using the delta method). Professor Ala Szczepura assisted by arranging meetings to help manage the project, and being a constant source of encouragement and support and proof reading final chapters.

With respect to the analysis in Chapter 5 of the thesis which I conducted and wrote up, Dr Higgins looked through a draft and made some small changes and small additions to the text. He also suggested other key references for inclusion in the associated paper that has been submitted to BMC Nephrology, and he is assisting me with final revisions (due back by 31st May 2012) prior to publication. Professor Leech gave me feedback on the econometric notation used. The authorship of the associated paper will be: Mr Michael D. Clark, Dr Anil K. Gumber, Professor Dennis Leech, Dr Domenico Moro, Professor Ala K. Szczepura, Mr Nicholas West, and Dr Robert M. Higgins. Much of the material in Chapter 6 of the thesis has already been published in the journal Diversity in Health, and the authorship for that paper was the same as for the paper submitted to BMC Nephrology. I analyzed the data and wrote the chapter, and published paper. Some minor changes were suggested by co-authors and peer reviewers, which I took on board.

In relation to the data analysis in Chapter 7, I had support from Dr Domenico Moro who familiarized me with how to conduct Mixed Logit and LCM, and also programming using GLAMMs in STATA. Although, I found learning how to conduct Mixed Logit straightforward, I did not fully grasp how to conduct Latent Class Modelling using GLAMMs the first time around. Therefore the LCM data analysis was conducted by me and him, under the direction of Dr Moro. That said I will shortly have to use LCM again to analyze data for my new renal dialysis DCE. I am therefore confident that having worked on LCM once again alongside Dr Moro, I will reach the point at which in future I could conduct such LCM analysis unaided. When it came to the written analysis contained in Chapter 7 I decided which econometric models to include, interpreted results, and wrote the chapter. A couple of helpful comments were made when I presented an early draft of analysis in Chapter 7 at the Health Economics Study Group in Bangor last year, which I took on board. Moreover, Professor Szczepura proof read the final chapter.

It was my idea to include both Eq-5d and VAS in the patient renal transplant questionnaire (see Appendix E) because I wanted to explore possible links between quality of life and DCE preferences. I also had the idea of exploring whether there might be a link between the perspective patients adopted when answering the questionnaire (i.e. altruistic or otherwise) and preferences (Chapter 8). I conducted all the data analysis for Chapter 8, and wrote the chapter. The material in that chapter was also proof read by Professor Szczepura. Also, Michael Alexander provided clerical assistance.

References.

Julian S., Naftalin N.J., & Clark M.D, et al. (2008). An integrated care pathway for menorrhagia across the primary-secondary interface: patients' experience, clinical outcomes, and service utilization. *Quality and Safety in Health Care.* 16, 110-115.

Street D.J, Burgess L., & Louviere J. (2005). Quick and easy choice sets: constructing optimal and nearly optimal stated choice experiments, *International Journal of Research Marketing*, 22, 459-470.

Introduction to the thesis.

Chapter 1 of the thesis is divided into 5 sections. Section A begins by defining a Discrete Choice Experiment (DCE), and outlines the key stages involved in conducting a DCE. Sections B and C outline theories underpinning DCE analysis. Section B outlines the characteristics theory of demand, whilst section C, explains random utility theory (RUT), compensating variation (CV), marginal willingness to pay (MWTP), and willingness to pay (WTP) analysis. Section D of the thesis provides a review of the DCE literature. Section E outlines the research questions addressed in the thesis including calculating WTP and hypothetical bias; the description of the cost attribute; preference heterogeneity; and altruism.

Chapter 2 shows how DCEs can be used to calculate WTP, using a DCE relating to Deep Vein Thrombosis (DVT). The regression results are used to show how MWTP can be calculated, and to construct value functions to predict the probability of uptake for service options, also welfare changes are calculated using a CV formula.

Chapter 3 uses data from a DCE applied to Menstrual disorder and Gynaecology patients. It evaluates an experimental method I developed to establish whether respondents might fail to factor in the monetary attribute into their DCE decision making, leading to hypothetical bias. All the DCE questionnaires for this data analysis had a monetary descriptor described as ‘cost to you.’ Using interaction dummy variables I then established whether estimated WTP differed between respondents who claimed not to take the monetary attribute into account in their decision making, and others who said they did or sometimes took it into account.

Chapter 4 applies essentially the same DCE design but only analyses data from Gynaecology patients. The scenarios are the same as those used for the DCE in chapter 3, except that there are 3 different descriptors for the monetary attribute. I tested for statistically significant differences in MWTP between DCE questionnaires

using each of the descriptors of the monetary attribute. I also repeated the test conducted in chapter 3 (for statistically significant differences in MWTP according to whether respondents took the monetary attribute into account).

Chapters 4-8 all use data obtained from a DCE relating to preferences for different allocation criteria for allocating kidneys for transplantation. Chapters 5 and 6 look at preference heterogeneity which is observable using interaction dummy variables (the issue of unobserved preference heterogeneity is considered in chapter 7). Chapter 5 establishes how marginal rates of substitution (MRS) differ between different respondent groups including renal patients, healthcare professionals, live donors / relatives of deceased donors, carers, and ethnic minority versus non-ethnic minority patients. Chapter 6 establishes how MRS differs between non-white ethnic minority patients versus other patients; South Asian patients versus other patients; and according to respondent gender. Chapter 7 of the thesis compares results from models which do not cater for unobserved preference heterogeneity, with results from models which do. Initially 2 basic models which do not cater for preference heterogeneity at all (because they do not include dummy variables) are applied including random effects logit and conditional logit. Then models catering for unobserved preference heterogeneity including Mixed Logit and a Latent Class Model (LCM) are used. Finally there is an analysis involving the application of conditional logit with interaction dummy variables. Chapter 8 of the thesis explores how preferences might differ according to how altruistic respondents are. It establishes how respondent preferences differ according to respondent self-disclosed perspective when answering DCEs. In other words whether they claimed to answer the DCE in terms of what would be best for me; what would be best for me and others; or what is best for others.

Finally chapter 9 involves a discussion of the findings emerging from the thesis, and draws conclusions about the merits of material contained in it.

Chapter 1: Explaining the nature of Discrete Choice Experiments (DCEs), and their foundations in economic theory, a review of DCE literature in health, and setting out the key issues examined in the thesis.

Introduction.

In section A of this chapter I begin by defining a DCE and then explaining the key stages involved in conducting one. In sections B and C the theoretical foundations of DCE analysis are explained. In section B the characteristics theory of demand is outlined, and details provided of how the theory can be used to allow for the assumption that chacteristics of goods or services can be valued in utility terms separately as with DCEs. In section C other theories underpinning DCE analysis are explained including random utility theory (RUT), Compensating Variation (CV), Marginal Willingness to Pay (MWTP) and Willingness to Pay (WTP).

Section D of this chapter outlines the key findings of 2 previously published reviews of the healthcare DCE literature relating to the period 1990-2008. I then updated these reviews by adding in the details of a further 96 papers published between 2009-2011. In section E of the chapter having conducted this literature review in section D, I then go on to outline 4 key methodological issues that the thesis will address in subsequent chapters.

A.1. What is a DCE?

Discrete choice experiments (DCEs) like Conjoint Analysis involves the application of a stated preference technique in order to establish respondent's valuation of attributes / characteristics

of a good or service or health state. However, unlike some Conjoint Analyses which are incompatible with random utility theory, DCEs are underpinned at a theoretical level by both Lancaster's characteristics theory of demand, and random utility theory (RUT) (see sections B and C of this chapter). DCEs therefore involve an attempt to value key defined attributes / characteristics of a good service, or health state, in a manner which is compatible with RUT.

A.1.1. key stages involved in conducting a DCE.

A number of key stages are associated with conducting a rigorous DCE.

1) Identifying a DCE application: Inevitably before embarking upon a DCE the suitability of a particular application has to be assessed. This should be an application in which underlying utility can in theory be separable (i.e. it should be possible to apply DCEs in order to establish how utility might be related to the levels of attributes /characteristics of a good, service, or health state, etc). Note although more basic DCE designs assume complete separability of utility, more elaborate designs are also possible in which potential interactions between attributes (i.e. whereby the utility value of levels of a given attribute is related to the levels of another or other attributes included a DCE) are explored. DCEs have been applied to a wide range of health related applications (for further details see sections D3 and D7.1.), and have involved the presentation of a wide range of summary outcome measures (see section D7.2).

2) Identifying DCE attributes and levels: A variety of strategies can be undertaken to identify suitable attributes to include in a DCE design (Coast and Horrocks 2007). Very often the first step involves conducting literature searches using appropriate databases and then

reviewing any literature obtained to establish whether it sheds some light upon what the attributes and levels might be. Sometimes expert opinion might be consulted. For example if it's a DCE for a health application, the advice of healthcare professionals who understand the issues surrounding the proposed application might be consulted in order to help inform an initial selection of attributes and levels for the DCE perhaps through interviews. Interviews or focus groups can also be conducted with individuals from the respondent group(s) who will ultimately be asked to complete a DCE in order to establish which attributes and levels might be valued by those who complete a DCE. Sometimes as (Coast and Horrocks 2007) point out patient surveys have been used to inform attribute and level selection. On other occasions suggestions have emanated from expert reviews (Hall, Kenny et al. 2002), or from the findings of clinical trials (Bryan, Buxton et al. 1998); (Ryan and Hughes 1997)).

3) Piloting the DCE: Although some DCEs have been undertaken without undertaking a pilot DCE first, the benefits of initial piloting and analysis of pilot data econometrically are considerable (a pilot DCE will need to be designed appropriately like a final DCE – see (4) below). More to the point if thorough qualitative analysis is not conducted to develop the attributes and levels then "lack of evidence of rigour casts doubts on the thoroughness of this work and thus the value of resulting DCMs" (Coast and Horrocks 2007). Personally, whenever I now embark upon a DCE, I ensure that I include a thorough pilot analysis, to obtain a mixture of both qualitative and quantitative information. Qualitative information can be obtained via either one-to-one interviews (Coast and Horrocks 2007), or via group discussions in focus groups / citizens juries. If one-to-one interviews are conducted during piloting of questionnaires, then 'think-aloud' exercises can be used (Ryan, Watson et al. 2009) to establish what meaning respondents attach to information contained in questionnaire

preambles; to the experimental pilot questionnaires attributes and levels; and also to other questions which might need to be posed relating to respondent characteristics, etc.

If a DCE involves the inclusion of a monetary attribute then particular attention needs to be paid to its appropriate description (in the interests of incentive compatibility). Moreover, piloting is needed to establish appropriate levels for the monetary attribute. The choice of monetary descriptor may be more straightforward, and the task of establishing appropriate levels for the monetary attribute simpler, if a DCE is applied in a context in which respondents currently have to pay for the healthcare provision in question. However, in many cases DCEs are applied to place a monetary valuation (using Willingness to Pay analysis [WTP] or Willingness to Accept [WTA] analysis) upon a good service or form of healthcare provision which is currently free at the point of use (or would be if it were to be made available). In these circumstances particular attention needs to be paid to ensure that the descriptor for the monetary attribute is as 'incentive compatible' as possible (i.e. the monetary attribute is phrased in such a manner that respondents are most likely to divulge their true trade-off between money and the other attributes, thereby providing an accurate indication of WTP or WTA). Also there is a need to ensure that an appropriate range is specified for the levels of the monetary attribute. This is because the range specified for the monetary attribute may affect estimated WTP or WTA. Moreover, if the range specified for the monetary attribute is not appropriate this can result in the monetary attribute appearing to be insignificant in econometric results (when in reality respondents would have valued changes in the monetary attributes, had changes been expressed over a more appropriate monetary range).

Unfortunately there is no definitive 'correct' method which can be applied to establish an appropriate range for the monetary attribute. However, the use of mainstream WTP

techniques including open-ended WTP / WTA analysis, or perhaps more appropriately using a series of payment cards to establish the ceiling level for the range for WTP / WTA, are approaches which are sometimes adopted.

During piloting it is also possible to extract information from respondents about whether they consider the range of attributes and levels included in the pilot DCE is appropriate, and also to invite respondents to suggest other possible attributes (or modifications to existing attributes) or changes to attribute levels. Respondents can also be invited (if they are filling in a pilot DCE questionnaire) to identify attributes currently specified in the pilot questionnaire, which they consider should either be excluded or amended for the purposes of the final DCE design. It is also possible to include attribute ranking exercises during piloting. These involve asking pilot respondents to rank the attributes already included in a DCE (plus others they suggest or other pilot respondents have suggested) in order of importance to them.

Whilst attribute ranking exercises might be of some value, one limitation of the approach is that respondents valuation of attributes might be a function of how much they value the stated levels for attributes (so that if specified attribute levels were changed their rankings might). Nonetheless, as long as the potential limitations of ranking exercises are recognised, information from attribute ranking exercises might provide an additional valuable source of information. Alternatively, other more qualitative iterative approaches can be adopted (Coast and Horrocks 2007)) in the interests of identifying and correctly phrasing DCE attributes and levels.

If enough respondents complete a pilot DCE questionnaire then the data can be analysed using an appropriate econometric technique. An analysis of the pilot data can provide

information about whether attributes (and sometimes levels of attributes) prove significant. Also if the pilot DCE questionnaire is framed to gather information about possible interaction effects between attributes, then it is possible to establish whether these interaction terms are significant in a pilot design, before catering for interactions (or perhaps the lack of interactions) in the final DCE design.

The quantitative information from analysis of pilot DCE data can be used alongside qualitative information, and perhaps quantitative information from pilot attribute ranking exercises to inform the specification of appropriate attributes and levels for inclusion in the final DCE analysis. In short conducting a rigorous pilot DCE analysis is essential in order to ensure that the final DCE design incorporates the most appropriate DCE attributes and levels. Moreover, the pilot exercise can be used to ensure interactions between attributes are appropriately catered for in the final DCE design. Also, qualitative feedback can help to ensure that attributes and levels are explained in a straightforward manner to DCE respondents. The pilot exercise can also be used to ensure that the range of additional socio-economic and other respondent information requested when a DCE is finally applied is appropriate, and that questions are framed in a manner which respondents can understand.

4) Designing a DCE: Careful attention is required in DCE design. If the design is for a pilot DCE, then the quality of information feeding into the DCE design (about attributes and levels for inclusion in the design) will not generally be as refined as for a final DCE. Nonetheless, even for pilot DCEs, it is usually possible to obtain information from the literature and from interviews with a small number of healthcare professionals in order to inform attributes and level selection for the pilot questionnaire.

If a pilot DCE has already been applied then the findings from quantitative analysis of pilot DCE data, and attribute ranking exercises (if they are conducted) alongside findings from qualitative analysis including interviews (sometimes involving 'think-aloud exercises) or focus groups (or other sources of information noted in 1 above) can be used to help inform the final selection of attributes and levels for the final DCE design. Researchers should also seek any information they can obtain during piloting, about whether respondents' valuation of attributes and levels overlap in utility space, and therefore whether two, or three way interactions ought to be incorporated in DCE designs.

One key consideration in the design of a DCE is to work out how many options respondents should be allowed to choose from per scenario.

A forced choice of one of two options may be useful if only 2 options are necessary and you want to avoid respondents having a 'don't know' option because they may be tempted to tick that box in preference to carefully choosing between the other 2 options (i.e. in cases for which the difference in utility between the 2 options is not that great and careful thought is required). However a forced choice approach has the disadvantage that respondents can only register their true 'indifference' between 2 options by not answering the question at all, rather than ticking a 'would not choose either' option. Therefore, even if there are only 2 options to choose between, some DCEs designs allow respondents to register a third 'don't want to choose' type of response.

In some contexts of course it may be applicable for respondents to face a choice between more than 2 options (which requires the use of multinomial models for data analysis

purposes, rather than binary dependent models which can be used if respondents only choose between 2 options).

DCE designs can be tailored to provide for the presentation of different numbers of options i.e. 2, 3 or more to options. Therefore, working out how many options respondents should face and how these ought to be presented is an important aspect of the DCE design, and an issue that needs to be addressed ideally during the pilot stage of DCE design.

For both pilot and final DCEs, if the number of attributes and levels is very small then it may be possible to apply a full factorial DCE design (i.e. use a design template which encompasses the full possible range of each attribute's levels). In contrast fractional factorial designs attempt to infer utility from a sub-set of scenarios which are comprehensive enough to enable utility to be inferred from a more limited range of choice scenarios that are presented to respondents. Most DCEs deploy fractional factorial designs because they involve specifying multiple attributes with multiple levels. As such, they would require the presentation of too many attributes and levels to respondents if a full, rather than fractional, factorial design was deployed.

By definition full factorial designs simply involve presenting respondents with all the attributes and all the different combination of attribute levels that may potentially arise, given the number of attribute and levels included in the design. So applying such designs is usually impractical unless the number of attributes and levels are both very low. If a fractional factorial design is applied then different approaches can be used to create an appropriate design (see section D.4). The standard approach for linear models is to adopt an orthogonal design (in which variations of attributes and alternatives are uncorrelated in all choice sets)

and then integrate them into choice designs (Amaya-Amaya M, Gerard et al. 2008). The approaches adopted usually involve the use of fractional factorial designs from catalogues or from the internet, or use of software packages (such as SPEED, SPSS, SAS, or other packages). Other approaches involve the use of “foldover” or “foldover with random pairing” type designs (see section D.4 for further details), or adopt statistically efficient designs (once again see section D.4 for further details), and the computer package SAS can be used to generate a D-optimal design.

D-optimal designs are allowed to deviate a little from orthogonal designs because they permit some limited correlations between attributes. This is in the interest of allowing ‘efficient’ estimation using smaller sample sizes. The package SAS (which uses a D-optimal approach) like some other packages is amenable to generating design templates for DCEs in which choice options are ‘labelled.’ Labelled choice DCEs are applicable if the choice between different options is not just a function of differences in attributes, but also of mode of choice. An example of this is the DCE I am currently undertaking (mentioned above), relating to healthcare professional, carer and patient preferences for different types of kidney dialysis. With kidney dialysis, some attributes of dialysis must assume certain attribute levels conditional upon the mode of dialysis adopted. For example with peritoneal dialysis, a small operation to the stomach is required (so that patients are dialysed via the stomach), whereas with haemodialysis blood is usually obtained and replaced via a fistula to the arm or neck. Therefore in such circumstances, in which the characteristics or attributes of provision are determined by mode of delivery, the nature of choice can only be appropriately presented using a labelled choice DCE approach. With labelled choice DCEs it is possible to value individual characteristics / attributes alongside a label which is also valued in utility terms (which may relate to mode of delivery, brand, name of supplier, etc).

5) Analysing DCE data: DCEs by definition must be compatible with random utility theory (RUT). This means that the data need to be analysed using an econometric model which is compatible with RUT. If only two choices (with no opt-out) are presented to a respondent (and preferences are not heterogeneous), then either Random Effects Logit or Random Effects Probit are probably the most appropriate models to adopt to analyse the DCE data econometrically.

If respondents are presented with more than 2 options, or 2 options plus an opt-out then the data can be analysed using a multinomial model such as multinomial logit (MNL). It should be noted that as discussed in section D.5 of this chapter, the MNL model can be applied if 3 main simplifying assumptions hold. MNL models are applicable if the following can reasonably be assumed:

(i) Independence of Irrelevant Alternatives (IIA).
(ii) Error terms are independently and identically distributed (IID) across observations.
(iii) No taste heterogeneity.

However, as discussed in section D.5, other models can be adopted if it is thought that any of these assumptions do not hold. For example, there are models which relax the IIA assumptions by allowing for more flexible error distributions i.e. Nested Logit and multinomial probit models. There are also heteroskedatic models which relax the IID assumption. Furthermore, there are models which cater for preference heterogeneity such as Mixed Logit (MXL) and Latent Class Models (LCM).

6) Presentation of DCE results: In section D of this chapter, I provide an overview and review of the DCE literature in the health field. In section D.7.1, I indicate the range of previous DCE health applications, and in section D7.2. I show how the output of DCEs has been presented in these analyses. When conducting DCEs it is logical to present results in terms of one or more key outcome measures.

These might involve expressing the value of attributes in terms of per willingness to pay unit; per time unit (perhaps waiting time); per risk unit; in terms of a monetary welfare measure; in terms of a utility score; in terms of odds ratios; or in terms of probability scores; or other summary outcome measures as outlined in section D.7.2.

The choice of appropriate key outcome measure(s) will depend upon the nature of the DCE application, and also the range of attributes specified in the DCE. The outcome measure(s) used should be tailored to the actual research question that needs to be addressed. So for example measures such as per WTP unit or a monetary welfare measure may be particularly appropriate outcome measures to include in a DCE analysis if researchers want to conduct a Cost-Benefit Analysis (CBA).

A measure such as 'per time unit' (which could be framed as waiting time) might be particularly appropriate in contexts where healthcare is free at the point of use (which might render summary outcome measures such as Willingness to Pay [WTP] subject to hypothetical bias). Because it has been argued that state funded healthcare systems often ration through the use of queues, a summary outcome measure such as willingness to wait (assuming a DCE includes an attribute relating to waiting time) may be more appropriate in such contexts. Summary outcome measures relating to risk units might be particularly appropriate if DCEs

are applied in contexts such as screening, diagnosis, or medical decision making, when the decisions to be addressed relates to how much respondents value increasing or reduced levels of risk. Utility scores can be used as summary outcome measures, but they may have the disadvantage that respondents may not be able to comprehend the value of the particular utility scores presented. Odds ratios may be useful if the information is required in relation to, for example, adoption rates of drugs or interventions. Equally, presenting results in terms of probability scores may be of value if DCEs are applied in contexts in which establishing relative probabilities is the main issue that a DCE needs to address.

When writing up DCE results it is good practice to provide a comprehensive account of the key stages that researchers have followed (i.e. steps 1-6 above). Moreover, when conducting DCEs it is common practice to incorporate some validity checks (see section D.6 of this chapter). Details of any validity checks undertaken, as well as of qualitative methods used to enhance DCE process and results, should be included when presenting DCE study results.

B. The theoretical underpinnings of Discrete Choice Experiments (DCEs) – The Characteristics Theory of Demand).

B1. The characteristics theory of demand.

(Lancaster 1966) provided a thorough characteristics theory of demand based on a critique of conventional consumer theory. He argues that “those properties that make a diamond quite obviously something different from a loaf of bread have been omitted from the theory” so that “the only property which the theory can build on is the property shared by all goods, which is simply that they are goods.” (Lancaster 1966) In other words he is saying that with conventional demand theory, demand is for goods or services, but that the theory makes no reference to their inherent characteristics. His ‘new approach’ therefore relates utility and demand to inherent characteristics. This is why this theory can provide the theoretical foundations for the use of DCEs, because DCEs relate utility to the inherent characteristics of goods or services.

Lancaster argued that conventional theory does not adequately deal with differences in goods, so it is ill equipped to deal with newly emerging differentiated goods, or by implication newly emerging services. One of the advantages of applying Discrete Choice Experiments is that you can value differentiated goods and services in terms of their differences in characteristics, and thereby relate differences in utility directly to differences in characteristics (referred to as the ‘attributes’ of a DCE).

B1.1.The characteristics theory of demand, and predicting demand.

With conventional theory, shifts in demand are often interpreted as changes in taste, a notion dismissed by Lancaster as 'non-operational' as "there is no way of predicting the relationship between *preference before and after the change*." (Lancaster 1966)[1] With Lancaster's framework though, since it is characteristics which are valued, goods or services which provide different configurations of characteristics (i.e. newly emerging goods), can be valued in terms of the value of their new inherent characteristics. Moreover, Discrete Choice Experiments (and Conjoint Analysis more generally) have been used by market researchers precisely for this purpose. Even within the narrow context of healthcare, DCEs have been used to predict demand for healthcare services or newly emerging healthcare services (Szeinbach, Mason et al. 1990); (Chakraborty, Ettenson et al. 1994); (Stensrud, Sylvestre et al. 1997); (Gates, McDaniel et al. 2000); (Payne and Elliott 2005).

B1.2.Substitution effects.

Lancaster's framework places the emphasis upon how much consumers value characteristics (the 'demand side'), but with the 'supply side' having an impact upon prices and what is provided to the market primarily as a result of the efficiency 'substitution effect'. This effect arises because consumers may change good collections as a result of compensated relative price changes, to obtain the same characteristics collection in a more efficient manner. There is also a private 'substitution effect', almost akin to the substitution effect with conventional demand theory (although operating at the level of characteristics rather than goods).

[1] My italics.

B1.3. Consumption activities.

Lancaster believes that consumption is an activity in which “goods, singly or in combination” (Lancaster 1966) are inputs into consumption activity, but that the desired output is a collection of characteristics. Services or goods provide the basis of “different joint outputs” (i.e. different bundles of characteristics). Therefore consumption activities generate joint outputs. Valuation of characteristics means “characteristics possessed by a good or combination of goods are the same for all consumers and, given units of measurement, are in the same quantities, so that the personal element in consumer choice arises in the choice between collections of characteristics only, not in the allocation of characteristics to goods.” (Lancaster 1966)

B 1.4. Lancaster’s theory of demand and utility?

The key elements of his theory are as follows:

1) Relating goods to consumption activities.

Goods or collections of goods are regarded as consumption activities to which a scalar (here k) defining a level of activity can be ascribed. The good in itself does not provide utility; it is the characteristics goods possess that confer utility. Lancaster allows for the fact that “Goods from different intrinsic commodity groups can be regarded as intrinsically unrelated, goods from the same commodity groups as intrinsically related.” (Lancaster 1966)

To simplify, goods are said to be consumed in linear activity, and objectively so that:

$$x_j = \sum_k a_{jk}\, y_k \qquad \text{(equation 1)}$$

yk: The level of consumption activity k

x_j: The j_{th} commodity.

a_{jk}: Determined by the intrinsic properties of goods themselves and possibly societies' technological knowledge.

Equation (1) therefore relates commodities to available goods. Equivalently, in matrix notation we have:

$$x = Ay \qquad \text{(equation 2)}$$

y: Goods

x: The commodity

A: The matrix relating the commodity to 'y' goods.

2) Relationship between activities and fixed vectors of characteristics.

Lancaster claims (and presumably for simplification) that consumption activities generate a fixed vector of characteristics (again via a linear relationship):

$z_i = \sum_k b_{ik} y_k$ (equation 3)

$z_{i:}$ Is the amount of the i^{th} characteristic.

b_{ik}: Relates goods consumed in a given activity (y_k's) via the b_{ik}'s to the given characteristics z_i's (B in matrix notation below), and in a sense indicates the sum of characteristic 'i', a good y_k is able to generate (coefficients b_{ik} are objectively determined) for some units of z_i.

Equivalently in matrix notation we have:

$z = By$ (equation 4)

3) Individuals have an ordinal utility function in characteristics space.

The function is U (z), and it is assumed to possess the convexity properties of a standard utility function.

B1.5. Key assumptions of the model.

In this model the z vectors define the direct ingredients of preferences i.e. characteristics (and thus welfare), which are related to the 'y' commodity goods; whilst the 'x' vector relates the x goods to the y commodity goods (i.e. the collections of goods to those available). Thus the vector x representing a consumer's relationship with the rest of the economy is not direct and one-to-one, as in the conventional model, but indirect, via the activity vector y - i.e. contingent upon the nature of the commodity goods (which of course can be regarded as different

bundles of characteristics). In effect, this means that different bundles of characteristics impact on demand for goods (the x's), contingent upon the bundles of characteristics intrinsic to the y goods.

In contrast to conventional consumer theory the "simple question asked (in principle) in the traditional analysis – does a particular consumer prefer collection x_1 or collection x_2 – no longer has a direct answer, although the question, does he prefer characteristics collection z_1 or z_2, does have such an answer." (Lancaster 1966). This therefore can imply a model of the following form:

Maximise $U(z)$ Subject to $px \leq k$

With $z = By$; $x = Ay$ and $z, x \geq 0$.

(equation 5)

Thus the maximand relates characteristics to utility (defined in C space – characteristics space), with a budget constraint based on the price of goods (in G-space), and a transformation relating the z characteristics (C-space) to the x goods (G-space), via B (which defines the consumption technology). Whilst the relationship between those x goods demanded, and the characteristics bundles available is determined by $x = Ay$.

To maintain this transformation from C to G-space, Lancaster claims it is necessary to assume that convex sets in G-space transform into convex sets in C-space, but points out that the inverse transformation between C-space and G-space may not exist (i.e. goods with a certain mapping of characteristics may not exist). Also non-negativity constraints are necessary.

For the most part it is assumed that either characteristics exceed the number of goods, or goods exceed the number of characteristics. Thus the utility function and the budget constraint must therefore be related in the same space. Either the utility function can be put in G-space [U(x)] [2], or the budget constraint can be expressed in C-space and related to U(z)[3]. With characteristics exceeding the number of goods (so that Bx = z contains more equations than variables x_i) we cannot find a goods vector giving rise to the arbitrarily specified characteristics vector z, thus maximise U(z) can be replaced by maximise U(x) and it will remain convex.

With goods exceeding the number of characteristics (the norm Lancaster claims in advanced societies), then the consumption technology has fewer equations than variables, thus for every characteristics vector there is more than one goods vector. Thus for every point in C-space the consumer can choose between different good vectors, implying a pure efficiency choice (i.e. delivering least cost bundles of characteristics through the choice of goods). Therefore for every characteristics vector, the consumer chooses the most efficient combination of goods to achieve an efficient bundle of characteristics (i.e. consumers are on the efficiency frontier).

B1.6. The efficiency frontier.

The efficiency choice is the solution to a canonical linear program:

[2] We can write U(z) = U (Bx) = u(x) to have a new utility function in terms of goods, but properties of u(x) depend on B(x).
[3] Again implies a more complex budget constraint than the conventional model.

Minimise: px Subject to: $Bx = z^*$; $x \geq 0$

(equation 6)

In other words expenditure is minimized to achieve a given level of characteristics (z) given goods x. By varying z*, the consumer given a constraint px = k can determine a characteristics frontier comprising all z such that the value of the above program is just equal to k (the level of consumption activity). Therefore there is a determinate goods vector associated with each point on the efficiency frontier. It is assumed that the consumption technology is objective, and the characteristics frontier is objective and is the same for all consumers facing the same budget constraint. There is also though in addition to the efficiency choice, a private choice i.e. a choice of which point on the efficiency frontier is preferred by a given consumer.

The characteristics frontier expands or contracts linearly (according to shifts in income), proportionately to an increase or decrease in income, so it has the same shape for all consumers who face the same prices. To demonstrate this simply and diagrammatically you can assume a situation in which there are 2 characteristics and 4 consumption activities. The horizontal and vertical axis respectively relate to 2 different characteristics z_1 and z_2. The equations below relate to the activities-characteristics portion of the consumption technology:

$z_1 = b_{11}y_1 + b_{12}y_2 + b_{13}y_3 + b_{14}y_4$ (equation 7)

$z_2 = b_{21}y_1 + b_{22}y_2 + b_{23}y_3 + b_{24}y_4$ (equation 8)

If there was only activity 1 then all goods purchased must be on ray 1 which is defined by b_{11} / b_{21}. Equivalently if there was only activity 2 the ray would be defined as b_{12} / b_{22}; if there was

only activity 3 it would be defined by b_{13} / b_{23}; and with only activity 4 it would be defined by b_{14}/b_{24}. There is a point on each of those rays at which given prevailing prices and income it is not possible to get further out from the origin (because of the income constraint). By joining up those points on the rays with straight lines between each ray and its nearest adjacent ray, you define the efficiency frontier.[4]

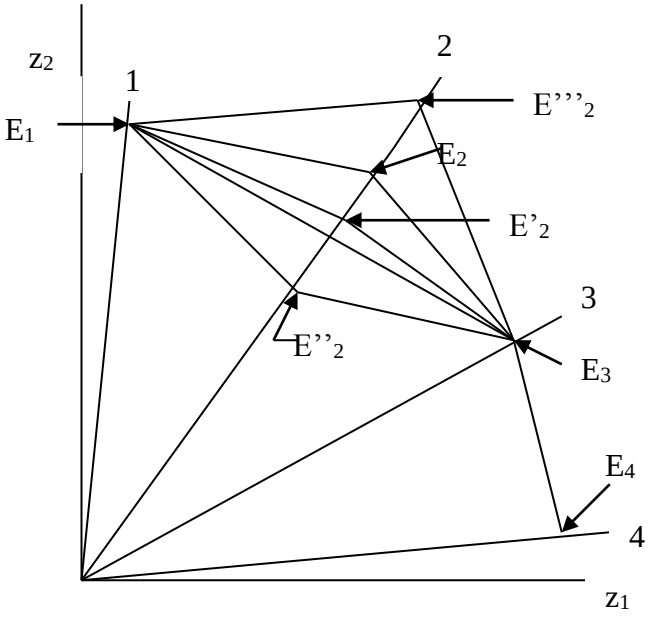


Figure 1: The efficiency frontier.

Figure 1 depicts the efficiency frontier, and then shows how it changes in response to a price change. We begin with an efficiency frontier defined by the lines between E_1, E_2, E_3, and E_4. Each of these points indicates the most (of combinations of characteristics z_1 and z_2) that with prevailing prices, can be obtained along a given ray for a given income. The effect of a price increase therefore in relation to one given good (i.e. good 2, along ray 2) is that for a given income, you can obtain less of good 2. This implies a shift inwards towards the origin along

[4] This of course assumes that they are on a ray which is either as far out or further out than a chord between the adjacent rays (in characteristics space).

ray 2. If this was sufficient to take you to E''_2 then with an efficiency substitution effect[5] this would take you inside the cord between E_1 and E_3, so E_2'' would no longer be on the efficiency frontier. Other bundles of characteristics provide more cost-effective means of generating characteristics, and good 2 would no longer be demanded (leading to an efficiency substitution effect). However if prices do not rise as much, and instead cause a movement along the ray to a point such as E'_2 then the efficiency frontier just shifts inwards such that its now defined by the points E_1, E'_2, E_3, and E_4. This is likely to result in a decline in demand for good 2 due to a private substitution effect (i.e. the fact that for some consumers their indifference curves are such that as the efficiency frontier shifts to E' they substitute towards other goods, as with conventional consumer theory). In contrast as you might expect a price decrease for good 2 implies that you can obtain more of good 2 for a given income and thus the efficiency frontier would shift out i.e. from E_1, E_2, E_3, and E_4 to E_1, E'''_2, E_3, and E_4.

Thus the analysis assumes that there is an efficiency substitution effect (switching effect) whereby if price changes are large enough there is a complete change from one activity to another. If you had similar but differentiated products (and a one to one relationship between goods and activities) then the efficiency substitution effect results in a complete shift from consumption of one good to another. If that is not the case and all goods are used in all activities, the efficiency substitution effect results in less consumption of a good whose price rises, but not complete disappearance from the market (i.e. like use of eggs in cakes).[6] Overall effects are reinforced by private substitution effects whereby consumers respond to changes in the shape of the efficiency frontier (resulting in some substitution of one good [characteristics bundle] for another).

[5] The efficiency substitution effect implies that consumers may change goods collections as a result of compensated relative price changes to obtain the same characteristics collection in a more efficient manner.
[6] A further requirement for the efficiency substitution effect is that the number of activities exceed the number of characteristics (otherwise switching of activities will not occur)

B.1.7. Conclusions (characteristics theory of demand).

Lancaster's characteristics theory of demand was the first serious theoretical attempt to construct a theory of demand which makes reference to both characteristics and goods space. It remains probably the most cited theory relating to the characteristics theory of demand. It is also often cited as providing the theoretical foundations for the use of Discrete Choice Experiments and some forms of conjoint analysis (CA). Moreover, the characteristics theory of demand also provides the theoretical underpinnings for analysis of revealed preference data, using hedonic pricing approaches.

C. Random Utility Theory (RUT), Compensating Variation (CV), Marginal Willingness to Pay (MWTP), and Willingness to Pay (WTP)

C.1 Random Utility Theory [RUT]

Lancaster's characteristics theory of demand provides the theoretical foundations for adopting an approach, which permits separable utility, and the related assumption that utility is related to characteristics (attributes) of goods and services.

Random Utility Theory [RUT] (McConnell 1995); (Mc Fadden 1999); (Louviere, Hensher et al. 2000) complements the characteristics theory of demand by providing a theoretical justification for the econometric estimation of demand. RUT models are useful because they provide for the assumption of deterministic utility maximizing behaviour and allow it to co-exist with apparent randomness across individuals which actual data exhibits. RUT assumes that consumers select the most attractive (utility maximising) options subject to constraints like time, and money.

RUT models assume that a respondent's preferences can be described by a utility function, and that respondents select an alternative conferring the greatest utility (Horowitz 1985); (Verbeek 2000). Therefore the utility of an alternative is represented by the sum of two components: a deterministic component accounting for the systematic effects of observed factors influencing choice, and a random component (hence the term 'random utility), accounting for the impact of unobserved factors (Horowitz 1985); (Louviere, Hensher et al. 2000). Therefore with the random utility framework, utility of each alternative is a linear

function of observed characteristics plus an additive unobservable error term (Verbeek), and thus using RUT utility (Louviere, Hensher et al. 2000) can be expressed as:

$U_i = V_i + \varepsilon_i$ (equation 9)

Here U_i is the latent measure of utility associated with option i, whilst V_i is the observable / systematic / explainable component of utility, and ε_i is the unobservable / unexplained component of utility. Given this underlying utility function, researchers want to model the probability that a randomly chosen consumer will select option i. This can be expressed (Louviere, Hensher et al. 2000) as:

$P(i \setminus A) = P[(V_i + \varepsilon_i) > \ldots > (V_j + \varepsilon_i) > \ldots > (V_j + \varepsilon_i)]$ for all j in A. (equation 10)

Here P (i \ A) is the probability that a consumer chooses action i from the set of all possible actions. These assumptions can then be used to underpin an econometric model which assumes that the effects of interest V_i are a function of observables (i.e. defined attributes in a choice experiment). Therefore V_i can be defined (Louviere, Hensher et al. 2000) as a linear function of observable parameters as:

$V_i = + \sum_k \beta_k X_{ki}$ (equation 11)

Here β is a K-element vector of parameters and X is an i by k vector, corresponding to observables describing actions that can be chosen. It is assumed that attributes will be selected for a choice experiment which correspond to the K-element vector of observable parameters affecting utility, and that respondents maximize their utility subject to constraints like time, and money (price).

Often, for DCEs (Louviere, Hensher et al. 2000) an econometric model will be assumed with the following underlying utility function:

$$U_i = + \sum_k \beta_k X_{ki} + \mu_i + \varepsilon_i \qquad \text{(equation 12)}$$

Such a model allows the error term in the regression to be decomposed to allow for multiple pairwise choice responses from each respondent (through the inclusion of μ_i) as well as other unobservable error terms (ε_i). Such a model can be estimated using a suitable random effects model.

C.2. Including a monetary attribute in a Random Utility Model.

If a price coefficient is included in the model, then it can be interpreted as the marginal utility of income so long as there is a binding budget constraint. Moreover, welfare measures can be derived based upon the concept of consumer surplus (Mcconnell 1995); (Mc Fadden 1974), which are equivalent to willingness to pay (WTP). In the McConnell paper consumer surplus (C) is expressed as:

$$C = \int_0 \prod_i (P)\, dP. \qquad \text{(equation 13)}$$

Here $\prod_i (P)$ is the demand curve indicating the probability of choosing alternative i, at price P, whilst dP indicates the change in price. The model assumes that C is the maximum amount an observer would expect to collect from an individual, and thus C is equivalent to WTP. McConnell argues that the consumer surplus measure obtained via calculation of WTP, can be used to establish Compensating Variation (CV), for a given price for alternative i.

CV is of course equivalent to the net revenue of a planner who has to compensate a consumer for a price change after it occurs, to return a consumer to her original utility level u^0. It would be negative if the planner would have to pay the consumer a positive level of compensation, because the price change makes her worse off (Mas-Colell, Whinston et al. 1995).

WTP measures, so long as they are unbiased and derived using a model consistent with RUT, are compatible with utility maximisation. However, for this to be the case an assumption of linearity with respect to the budget constraint is required. Typically, within health economics, estimates of WTP derived using DCEs take the form of estimates of Marginal Willingness To Pay (MWTP). These are usually obtained using the following formula (Lancsar and Savage 2004)):

$$MWTP_1 = (\partial IUF / \partial X_1) \ / (\partial IUF / \partial P) = \beta_1 / - \beta_p \qquad \text{(equation 14)}$$

Here IUF is the indirect utility function estimated using a DCE, and P is price, whilst X_1 indicates attribute 1 from the vector X. In effect therefore MWTP for a given change in attribute X_1 is normalized into monetary terms, by taking the impact upon indirect utility for the model overall as a result of a change in attribute X_1, and dividing through by the impact in indirect utility for a given change in price (usually defined in terms of 1 unit of a currency: £1, $1, or €1, etc). For attribute X_1 it is equivalent to the coefficient on X_1 (which is β_1) divided by the coefficient on the monetary attribute which is price (which is β_p). Improvements with respect to a change in X_1 should be associated with positive values of WTP, whilst undesirable changes in X_1 should be associated with negative values of WTP for a given change in X_1.

Alternatively if you want to assess a change in WTP arising from a change in all attribute levels of a product or service the following formula can be used (and commonly is) to quantify WTP:

$$WTP = \sum_{k} (\beta_k / -\beta_p) \cdot (\Delta X_k) \qquad \text{(equation 15)}$$

Here the subscript k represents the attributes describing the product or program. This method involves the unweighted summation of the product of MWTP multiplied by the change in levels across attributes. However, Lancsar & Savage assert that using this formulation is appropriate "only if the alternative is chosen with certainty." (Lancsar and Savage 2004), which will be the case (Ryan 2004) for 'state of the world' models.

D. A systematic review of the literature relating to Discrete Choice Experiments (DCEs) in healthcare.

D.1. Introduction.

The most comprehensive reviews of DCEs published to date are (Ryan and Gerard 2003); and (de Bekker-Grob, Ryan et al. 2012). In this chapter, some of the results of these earlier reviews (covering 1990 up to the end of 2008) are indicated. I also update and build upon them, by reviewing DCEs in health covering the period 2009 - 2011.

Both the previous reviews had inclusion criteria which sought to identify health related, discrete choice experiment studies written in the English language. The earlier review (Ryan and Gerard 2003) used the following sources to obtain potentially relevant literature: MEDLINE; EMBASE; HEALTHSTAR; Social Science Citation Index; PsychLIT; Econlit; and the Health Management Information Consortium database. The search terms used included 'conjoint', 'conjoint analysis', 'conjoint measurement', 'conjoint studies', 'conjoint choice experiments', 'part-worth utilities', 'functional measurement', 'paired comparisons', 'pairwise choices', 'discrete choice experiments' and 'stated preference.' The more recent review (de Bekker-Grob, Ryan et al. 2012), applied the same search terms, using Medline to identify studies, but involved what the authors describe as "a narrower search of databases" which was not fully described in the paper. I contacted the lead author (E-mail communication with Dr de-Bekker-Grob) and discovered that the only database used for the review was PUBMED

/ MEDLINE. The authors (de Bekker-Grob, Ryan et al. 2012) did point out in their paper that they "have since re-run a broader search for another purpose which revealed other studies but the broad findings of this review do not change"; so it would appear that narrowing the search to studies cited only in PUBMED / MEDLINE did not significantly impact upon the conclusions of their review.

The first review (Ryan and Gerard 2003) covered papers published over the period 1990-2000, whilst the latter focused on papers published between 2001 and 2008. The 2003 review involved literature searches for the period 1990-2000 and generated 919 potential references. However, the pool of potential studies was reduced to 328 once abstracts for conjoint analysis studies not grounded in random utility theory (RUT), revealed preference studies, and studies not written in English or without health economics applications were removed. After a second round of more careful abstract reading, a further 199 abstracts were removed from the review by the authors on the grounds that they were not original empirical studies.

Therefore 129 full papers were obtained, out of which 95 studies were rejected for a variety of other reasons. Thus, the first review covering the period 1990-2000 only identified 34 studies which met its inclusion criteria.

The second review was conducted by de-Bekker-Grob (de Bekker-Grob, Ryan et al. 2012)), in collaboration with both authors of the original review, and therefore represented an updating and continuation of the earlier review (Ryan and Gerard 2003). The more limited literature searches in this second review generated 682 possible references for the period 2001-2008, and after reading abstracts or full articles, the authors concluded that 121

references relating to 114 original studies met their inclusion criteria. The authors also noted that the number of applications of discrete choice experiments in healthcare had risen from a mean of 3 per year in 1990-2000 to a mean of 14 per year during 2001-2008.

D.2. Methods.

The present review builds upon these two earlier reviews. It focuses on the literature for the 3 year period of 2009 —2011, identified using the same search terms. Like the authors of the most recent review (de Bekker-Grob, Ryan et al. 2012), I restricted the searches to PUBMED / MEDLINE. I conducted the initial searches in September 2011,and updated these in March 2012 to ensure that all 2011 papers were identified and included in the review. Thus, the review covered a full 3 year period from the 1st January 2009 – 31st December 2011. The searches conducted in September 2011 generated 218 citations and searches in March 2012 identified a further 50 references. After examining abstracts or papers for these references, 96 studies were identified which met the inclusion criteria for review i.e. they related to Discrete Choice Experiments compatible with Random Utility Theory (RUT) and had been published in English speaking journals. Articles reporting Adaptive Conjoint Analysis methodology were excluded, since Esther de Bekker-Grob had informed me by e-mail that these had been excluded from their review (de Bekker-Grob, Ryan et al. 2012).

D.3. Overview of DCE applications.

Over the period 2009 – 2011 the number of papers relating to health related discrete choice experiments studies fulfilling the inclusion criteria was 98. However, these papers only related

to 96 [7] distinct studies, as 2 papers (Guimaraes, Marra et al. 2010; Boonen, Donkers et al. 2011) replicated analyses in other papers published during this period, so were removed. Therefore, the mean number of studies per year in this period has risen to 32.

The most recent review (de Bekker-Grob, Ryan et al. 2012) subdivided the papers reviewed by study objective (some papers appeared under 2 or more categories if they had multiple objectives). For the current review, I also sub-divided papers into the same broad categories. In table 1, a summary is provided of descriptors for papers included in the 3 reviews. In addition, in the 2001 to 2008 review papers were classified into a number of broad categories (see (de Bekker-Grob, Ryan et al. 2012)). Below, I indicate the number of papers for 2009 — 2011 falling into these same broad categories compared with 2001 to 2008 figures:

A. *Patient or consumer experience factors - 40 studies for 2001 to 2008; 16 studies for 2009 to 2011:* (Albada and Triemstra 2009; Boonen LHHM, Schut FT et al. 2009; Brown, Finkelstein et al. 2009; Clark MD, Gumber AK et al. 2009; Davison, Kromm et al. 2010; Deverill, Lancsar et al. 2010; Kiiskinen, Suominen-Taipale et al. 2010; Nieboer, Koolman et al. 2010; van der Pol and McKenzie 2010; van der Pol, Shiell et al. 2010; Darba, Restovic et al. 2011; Mentzakis, Stefanowska et al. 2011; Pereira, Mulligan et al. 2011; Potoglou, Burge et al. 2011; Waltzman, Scholz et al. 2011; Yi, Ryan et al. 2011).

B. *Valuing health outcomes - 8 studies for 2001 to 2008; 13 studies for 2009 to 2011:* (Bederman S, Mahomed NN et al. 2009; Goto, Takahashi et al. 2009; Johnson, Hauber et al. 2009; Ratcliffe, Brazier et al. 2009; Skjoldborg, Lauridsen et al. 2009; van Til, Stiggelbout et al. 2009; Witt, Scott et al. 2009; Koopmanschap, Stolk et al. 2010; Lancsar, Wildman et al.

[7] It is sometimes difficult to know whether analyses should be classed as separate studies or not. In the end we decided to treat the two analyses by van Helvourt-Postulart et al (2009) as 2 separate studies, because they related to 2 different areas of study. One of them focused itself with application of hierarchical information integration to DCEs; whilst the other was a comparison of DCEs and "barriers and facilitators in implementation research".

2011; Mentzakis, Ryan et al. 2011; Poulos, Yang et al. 2011; Ratcliffe, Couzner et al. 2011; Vroomen and Zweifel 2011).

C. *Investigating trade-offs between health outcomes and patient or consumer experience factors - 38 studies for 2001 to 2008; 51 studies for 2009 – 2011:* (Bunge EM, de Bekker Grob EW et al. 2009; Chan, Sahota et al. 2009; de Bekker-Grob, Essink-Bot et al. 2009; Eberth, Watson et al. 2009; Green and Gerard 2009; Guimaraes, Marra et al. 2009; Hauber, Mohamed et al. 2009; Howard and Salkeld 2009; Kruijshaar, Essink-Bot et al. 2009; Muhlbacher, Rudolph et al. 2009; Ozdemir, Johnson et al. 2009; Pavlova, Hendrix et al. 2009; Regier, Friedman et al. 2009; Scalone, Mantovani et al. 2009; Tinelli, Ryan et al. 2009; Bijlenga, Bonsel et al. 2010; de Bekker-Grob, Hofman et al. 2010; Essers, Dirksen et al. 2010; Essers, van Helvoort-Postulart et al. 2010; Hauber, Mohamed et al. 2010; Hol, de Bekker-Grob et al. 2010; Johnson, Ozdemir et al. 2010; Nayaradou, Berchi et al. 2010; Scuffham, Whitty et al. 2010; Torbica and Fattore 2010; van Dam, Hol et al. 2010; Wittink, Cary et al. 2010; Ahmed and Fincham 2011; Bogelund, Vilsboll et al. 2011; Brown, Pashos et al. 2011; Damen, de Bekker-Grob et al. 2011; Faggioli, Scalone et al. 2011; Goto, Takahashi et al. 2011; Hauber, Gonzalez et al. 2011; Lagarde, Smith Paintain et al. 2011; Laver 2011; Lloyd, Hodgkins et al. 2011; Lloyd, Nafees et al. 2011; Marti 2011; Mohamed, Epstein et al. 2011; Muhlbacher and Nubling 2011; Musters, de Bekker-Grob et al. 2011; Oteng, Marra et al. 2011; Scalone, Watson et al. 2011; Schwappach, Mulders et al. 2011; Scotland, McNamee et al. 2011; Sweeting, Whitty et al. 2011; Swinburn, Lloyd et al. 2011; Thrumurthy, Morris et al. 2011; van Empel, Dancet et al. 2011; Van Houtven, Johnson et al. 2011; Hong, Liu et al. 2011).

D. *Estimating utility weights within the QALY framework - 2 studies for 2001 to 2008; 2 studies for 2009 – 2011:* (Stolk, Oppe et al. 2010; Lancsar, Wildman et al. 2011).

E. *Job-choices - 6 studies for 2001 to 2008; 4 studies for 2009 to 2011:* (Blaauw, Erasmus et al. 2010; Grindrod, Marra et al. 2010; Gunther, Kurstein et al. 2010; Kolstad 2011).

F. *Developing priority setting frameworks - 6 studies for 2001 to 2008; 7 studies for 2009 – 2011:* (Albada and Triemstra 2009; Clark MD, Gumber AK et al. 2009; Green and Gerard 2009; Davison, Kromm et al. 2010; Koopmanschap, Stolk et al. 2010; Youngkong, Baltussen et al. 2010; Watson, Carnon et al. 2011).

G. *Health professional's preferences for treatment or screening options for patients - 17 studies for 2001 to 2008; 9 studies for 2009 – 2011:* (van Helvoort-Postulart, Dellaert et al. 2009; van Helvoort-Postulart, van der Weijden et al. 2009; Bhatt, Currie et al. 2010; Tsung-Tai C 2010; Wyatt, Batley et al. 2010; Faggioli, Scalone et al. 2011; Lagarde, Smith Paintain et al. 2011; van Empel, Dancet et al. 2011; Benjamin, Cotte et al. 2012).

H. *Other - 4 studies for 2001 to 2008; 5 studies for 2009 to 2011:* (Ratcliffe, Brazier et al. 2009; Tsung-Tai C 2010; Faggioli, Scalone et al. 2011; Fegert, Slawik et al. 2011; Whitty, Scuffham et al. 2011).

Each paper was read carefully and key data extracted in a systematic manner. This is presented in the following sections.

D.3.1. Number of DCE analyses per year.

de Bekker-Grob et al (de Bekker-Grob, Ryan et al. 2012) noted that the number of applications of discrete choice experiments in healthcare rose from a mean of 3 per year (1990 — 2000), to a mean of 14 per year (2001 - 2008). Our review for 2009 —2011 obtained papers (meeting the inclusion criteria) relating to 96 different healthcare related discrete

choice experiments studies. Overall 29 of these papers were published in 2009; 27 in 2010; and 40 in 2011. The average number of papers per year therefore rose to 32 (2009 — 2011) which is a very marked increase compared to the earlier periods.

D.3.2. DCE studies country of origin.

In the 2001 – 2008 review it was noted that the UK remained the major user of DCEs. However, the UK's dominance has been eroded considerably in these years. Table I shows that the proportion of studies emanating from the UK has continued to fall, from 20 / 34 (59%) in 1990-2000, to 55 / 114 (48%) in 2001 – 2008, to a low of 20 / 96 (21%) during 2009 – 2011. Moreover, the number of studies emanating from Australia has also fallen proportionately from 7 / 34 (18%) in 1990 – 2000, to 13 / 114 (11%) in 2001 – 2008, to 7 (7%) in 2009 – 2011.

Comparing 1990 – 2000 with 2001 – 2008, the proportion of studies coming from Canada rose slightly (1 / 34 (3%) in 1990 – 2000 to 6 / 114 (5%) in 2001 – 2008), as did the number of studies from other countries such as Denmark (rose from 0 to 5 / 114 (4%) in 2001 – 2008); the Netherlands (rose from 0 to 5 / 114 (4%) in 2001 – 2008); Germany (rose from 0 to 3 / 114 (3%) in 2001 – 2008), and 'Other' countries (rose from 0 to 13 / 114 (11%) in 2001 – 2008).

However, comparing 2001-2008 with the most recent 2009 – 2011 period, it is clear that an increasing proportion of studies now originate in places such as the USA (14 / 114 (12%) in 2001-2008 up to 16 / 96 (17%) in 2009 – 2011); Canada (6 / 114 (5%) in 2001 – 2008 up to 9 / 96 (9%) in 2009 – 2011); Denmark (5 / 114 (4%) in 2008 – 2011 up to 6 / 96 (6%) in 2009 –

2011); the Netherlands (5 / 114 (4%) in 2001 – 2008 up to 18 / 96 (19%) in 2009 – 2011); Germany (3 / 114 (3%) in 2001 – 2008 up to 6 / 96 (6%) in 2009 – 2011). There is also an increase in studies coming from 'Other' countries (13 / 114 (11%) in 2001 – 2008 up to 19 / 96 (20%) in 2009 – 2011) which reflects an increasing trend towards applying DCEs in an increasing range of different countries including high, middle and low income countries.

D.3.3 The number of attributes included in DCE studies.

Examining the number of attributes included in studies shown in table 1, it is clear that the trend towards fewer studies with only 2 – 3 attributes has continued. In 1990 – 2000 5 / 34 (15%) studies fell in this category, 15 / 114 (13%) during 2001 – 2008, and the figure with 2 – 3 attributes fell to 10 / 96 (10%) in 2009 – 2011. The proportion of studies with 4 -5 attributes rose from 10 / 34 (29%) in 1990 – 2000 to 50 / 114 (44%) in 2001 – 2008, but it fell back again to 29 / 96 (30%) in 2009 – 2011.

Closer examination indicates that the trend away from DCE studies with either 2 – 3 attributes (down from 13% to 10% between 2001 – 2008 and 2009 - 2011), or 4 to 5 attributes (down from 44% to 30% between 2001 – 2008 and 2009 - 2011), is partly attributable to a trend towards more DCEs with between 6 and 9 attributes. For example the proportion of studies with 6 attributes was stable at 9 / 34 (26%) in 1990 – 2000 and 30 / 114 (26%) in 2001 – 2008, but has risen marginally to 28 / 96 (29%) in 2009 – 2011. Moreover the proportion of studies with 7 – 9 attributes rose from 4 / 34 (12%) in 1990 – 2000, to 15 / 114 (13%) in 2001 – 2008, and then to 22 / 96 (23%) in 2009 – 2011.

The proportion of studies with 10 attributes has remained at 2% during both 2001 – 2008 (2 / 114) and 2009 – 2011 (2 / 96), falling from 2 / 34 (6%) in 1990 – 2000. The proportion of studies with more than 10 attributes fell from 4 / 34 (12%) in 1990 – 2008 to 2 / 114 (2%) in 2001 – 2008, but rose again slightly to 5 / 96 (5%) in 2009 – 2011.

D.3.4. Domains of DCE attributes.

The proportion of DCE studies including a monetary measure remained fairly constant over the period. In 1990 – 2010, it was 19 / 34 (56%), in 2009 – 2011, 61 / 114 (54%) and in 2009 – 2011 53 / 96 (55%). The proportion of DCE studies including a measure of risk rose in the most recent period, from 12 / 34 (35%) in 1990 – 2000 and 35 / 114 (31%) in 2001 – 2008, to 62 / 96 (65%) in 2009 – 2011 (the high figure for 2009 - 2011 may in part be due to a somewhat broad interpretation of what might be construed as an attribute relating to risk).

Finally, the proportion of studies with a health status domain showed a similar pattern, with 19 / 34 (56%) in 1990 – 2000, 61 / 114 (54%) during 2001 – 2008, and rising to 70 / 96 (73%) in 2009 – 2011. At the same time, the proportion of studies with attributes relating to other domains, increased from 3 / 34 (9%) in 1990 – 2000 to 17 / 114 (15%) in 2001 – 2008, and then to 39 / 96 (41%) for 2009 – 2011.

D.3.5. The number of questions posed by DCEs.

Between the 2001 – 2009 review and the 2009 – 2011 period, there has been a clear trend away from having 8 or fewer choices towards having 9 – 16 choices per DCE, as shown in table 1. The proportion of DCE studies posing 8 or less choices was 13 / 34 (38%) in 1990 -

2000, rising marginally to 45 / 114 (39%) in 2001 – 2008, and then falling back to 22 / 96 (23%) in 2009 – 2011. In contrast the proportion of studies with 9 – 16 choices was 18 / 34 (53%) in 1990 – 2000, fell to 43 / 114 (38%) during 2001 – 2008, and then rose to 53 / 96 (55%) during 2009 – 2011. The proportion of studies with more than 16 choices rose initially and then stabilized. It was 2 / 34 (6%) in 1990 – 2008, then 21 / 114 (18%) in 2001 – 2008, and 18 / 96 (19%) in 2009 – 2011. At the other end of the spectrum, the proportion of studies which did not report the number of choices that respondents had to face has remained relatively low; 1 / 34 (3%) in 1990 – 2000, 5 / 114 (4%) in 2001 – 2008, and rising marginally again to 5 / 96 (5%) in 2009 – 2011.

D.3.6. DCE survey administration.

In terms of 'Administration of the survey', table 1 indicates that there appears to be a trend since 1990 away from self-completed questionnaires. The proportion of self-completed questionnaires was 27 / 34 (79%) in 1990 – 2000, falling to 76 / 114 (67%) in 2001 – 2008, and then falling further to 58 / 96 (60%) in 2009 – 2011. The proportion of interviewer administered DCEs was 3 / 34 (9%) in 1990 – 2000, rose to 22 / 114 (19%) in 2001 – 2008, but then fell back marginally to 16 / 96 (17%) in 2009 – 2011.

Overall, there has been a general trend towards DCEs involving a computerized review over the last twenty years. During 1990 – 2000 3 / 34 (9%) of respondents answered DCEs which involved computerized review, in 2001 – 2008 the proportion increased to 13 / 114 (11%) and then rose sharply to 25 (26%) during 2009 – 2011. The proportion of DCEs not reporting details about survey administration was low, but rose slightly from 1 / 34 (3%) in 1990 – 2000, to 9 / 114 (8%) in 2001 – 2008, but then fell back again to 2 / 96 (2%) during 2009 – 2011.

Table I. Background information on DCE studies.

Item	Category	Baseline: 1990 – 2000	Period: 2001 – 2008	Period: 2009 – 2011
		N = 34	N = 114	N = 96
Country of origin	UK	20 (59%)	55 (48%)	20 (21%)
	USA	7 (21%)	14 (12%)	16 (17%)
	Australia	6 (18%)	13 (11%)	7 (7%)
	Canada	1 (3%)	6 (5%)	9 (9%)
	Denmark	0 (0%)	5 (4%)	6 (6%)
	Netherlands	0 (0%)	5 (4%)	18 (19%)
	Germany	0 (0%)	3 (3%)	6 (6%)
	Other	0 (0%)	13 (11%)	19 (20%)
Number of attributes	2 – 3	5 (15%)	15 (13%)	10 (10%)
	4 – 5	10 (29%)	50 (44%)	29 (30%)
	6	9 (26%)	30 (26%)	28 (29%)
	7 – 9	4 (12%)	15 (13%)	22 (23%)
	10	2 (6%)	2 (2%)	2 (2%)
	>10	4 (12%)	2 (2%)	5 (5%)
Attributes covered	Monetary measure	19 (56%)	61 (54%)	53 (55%)
	Time	25 (74%)	58 (51%)	71 (74%)
	Risk	12 (35%)	35 (31%)	62 (65%)
	Health Status domain	19 (56%)	62 (54%)	70 (73%)
	Health care	28 (82%)	79 (69%)	75 (78%)
	Other	3 (9%)	17 (15%)	39 (41%)
Number of choices per respondent	8 or less	13 (38%)	45 (39%)	22 (23%)
	9 – 16 choices	18 (53%)	43 (38%)	53 (55%)
	More than 16 choices	2 (6%)	21 (18%)	18 (19%)
	Not clearly reported	1 (3%)	5 (4%)	5 (5%)
Administration of survey	Self-completed questionnaire	27 (79%)	76 (67%)	58 (60%)
	Interviewer administered	3 (9%)	22 (19%)	16 (17%)
	Computerised review	3 (9%)	13 (11%)	25 (26%)
	Not reported	1 (3%)	9 (8%)	2 (2%)

D.4. DCE Experimental design and choice set construction.

After selecting appropriate attributes and levels, the DCE must adopt a suitable design. Full factorial designs can be deployed which present respondents with scenarios reflecting all the possible attributes and levels which might be included in a DCE. However, such designs may be too large for respondents to complete, which is why smaller fractional factorial designs are usually preferred.

The standard approach for linear models is to adopt an orthogonal design (these have the property that variations of attributes and alternatives are uncorrelated in all choice sets i.e. they are statistically independent of each other) and then integrate them into choice designs (Amaya-Amaya M, Gerard et al. 2008). Truly orthogonal designs also exhibit 'level balance' (de Bekker-Grob, Ryan et al. 2012) i.e. the levels of attributes each appear with equal frequency. As de-Bekker Grob et al (de Bekker-Grob, Ryan et al. 2012)) point out orthogonal designs can be sourced from certain design catalogues (Hahn and Shapiro 1966), some software packages (e.g. SPEED (Bradley 1991) or SPSS) or from websites (e.g. (Sloane 2009)). In the case of binary choice DCEs the profiles generated from orthogonal designs can be the choices (de Bekker-Grob, Ryan et al. 2012).

In some instances choice sets have been created from orthogonal arrays, and two main methods (Louviere, Hensher et al. 2000)) can be used (de Bekker-Grob, Ryan et al. 2012). One of these involves pairing choices with their 'foldover' which may refer to the mirror image of the original design (such that for example 16 profiles might be paired with their foldover, thereby generating 16 choices – for further details see footnote 2, in (de Bekker-Grob, Ryan et al. 2012). Alternatively the approach of 'foldover with random pairing' may be deployed (so

that the 16 profiles might be randomly paired with their foldover). As (de Bekker-Grob, Ryan et al. 2012) point out some designs are readily available which use these methods (Burgess and Street 2003), (Burgess and Street 2005); (Street and Burgess 2004); (Street and Burgess 2007a); (Street, Burgess et al. 2005), (Street, Burgess et al. 2008)) as well as help from experts (Street and Burgess 2007b).

More recently, as de-Bekker Grob et al (de Bekker-Grob, Ryan et al. 2012) point out, statistically efficient designs have been developed (Burgess and Street 2003), (Burgess and Street 2005); (Street, Burgess et al. 2005); and (Zwerina, Huber J et al. 2005). Numerous measures of statistical efficiency can be deployed but the most common is the D-efficient criterion. With D – optimal designs the determinant of the covariance matrix is minimised (de Bekker-Grob, Ryan et al. 2012). D-optimal designs deviate a little from orthogonal designs because they permit some limited correlations between attributes. This is in the interest of allowing 'efficient' estimation using smaller sample sizes (Amaya-Amaya M, Gerard et al. 2008). The computer package SAS, which is increasingly applied in order to produce DCE design templates, includes a D-optimal design algorithm.

When deriving efficient designs it is commonly assumed that the parameters are zero (de Bekker-Grob, Ryan et al. 2012). However as Amaya-Amaya, Gerard, and Ryan (Amaya-Amaya M, Gerard et al. 2008) point out deriving efficient designs may involve making prior assumptions about parameter estimates (this approach is exemplified by the work of Sandor and Wedel, (Sandor and Wedel 2001), (Sandor and Wedel 2002), and (Sandor and Wedel 2005)) and then generating a D-optimal design (the package SAS can be used to generate D-optimal designs). Prior assumptions about parameters may be informed by analysing pilot DCE questionnaire data econometrically, and then incorporating prior estimates of

parameters in order to increase statistical efficiency into designs generated using packages such as SAS or Ngene (de Bekker-Grob, Ryan et al. 2012).

D.4.1. Design type.

Table II provides an overview of changes in the design type adopted by researchers. This shows that the proportion of studies involving full factorial DCE designs fell from 4 / 34 (12%) in 1990 – 2000 to 0% in 2001 to 2008, and then rose again marginally to 4 / 96 (4%) during 2009 – 2011. In the period 1990 – 2000, 25 / 34 (74%) of studies were fractional factorial studies, a figure which rose to 114 / 114 100% in 2001 – 2008, but then fell back to 86% in 2009 – 2011. Overall, 5 / 34 (15%) of authors did not clearly report their 'Design type' in 1990 – 2000, a figure which fell to 0% in 2001 – 2008, but then rose again to 9 / 96 (9%) during 2009 – 2011.

Overall in 1990 – 2000 25 / 34 (74%) of DCE studies involved a 'main effects' design, and this proportion rose to 100 / 114 (89%) during 2001 – 2008, although the figure fell back to 68 / 96 (70%) in 2009 – 2011. So, as with the baseline review and 2001 – 2008 review, 'main effects' designs remain the dominant type of design in published DCE studies. In 1990 – 2000, 2 / 34 (6%) of study designs catered for interaction effects; this figure fell slightly to 6 / 114 (5%) in 2001 – 2008, before rising again to 11 (11%) in 2009 – 2011. I anticipate a continued trend towards increasing use of designs which cater for interactions, given that there is increased awareness of the need for designs to cater for interactions between attributes, where interaction effects might exist. In 1990 – 2000 it was suggested that for 4 / 34 (12%) of studies a design plan was not applicable, this figure fell back to 0% in 2001 – 2008, before rising very slightly to 1 / 96 (1%) in 2009 – 2011. However, the proportion of studies for which

design type was not clearly reported was 3 / 34 (9%) in 1990 – 2000, 0% in 2001 – 2008, and 16 / 96 (17%) in 2009 – 2011.

D.4.2. Use of software packages to design DCEs.

The use of a software package to design DCEs has fallen slightly. Overall 19 / 34 (56%) of designs were said to involve use of a software package in 1990 – 2000, falling slightly to 59 / 114 (52%) in 2001 – 2008, and then falling again to 44 / 96 (46%) in 2009 – 2011. There seems to have been a general trend away from using the package SPEED over the period, probably because it has become increasingly dated in nature, and new packages have become available. In 1990 – 2000 13 / 34 (38%) of studies involved the use of SPEED; this proportion fell to 22 / 114 (19%) in 2001 – 2008, and 6 / 96 (6%) in 2009 – 2011. Use of SPSS varied; in 1990 – 2000 2 / 34 (6%) of studies used SPSS, rising to 14 / 114 (12%) in 2001 – 2008, before falling back to 7 / 96 (7%) in 2009 – 2011.

The package SAS (which provides D-optimal designs) has become increasingly popular over the period. Recorded use rose from 0% in 1990 – 2000, to 14 / 114 (12%) in 2001 – 2008, and then to 18 / 96 (19%) in 2009 – 2011. Sawtooth was less popular, remaining relatively constant; 2 / 34 (6%) of authors in 1990 – 2000, 5 / 114 (4%) in 2001 – 2008, and 5 / 96 (5%) in 2009 – 2011. The use of ‘Other’ software was also low; 2 (6%) in 1990 – 2000, 0% in 2001 – 2008, and rising again to 7 / 96 (7%) in 2009 – 2011. A few papers included ‘No further details’ relating to ‘Design Source’; 0% cases in 1990 – 2000, 4 / 114 (4%) in 2001 – 2008, and 2 / 96 (2%) in 2009 – 2010.

D.4.3. Use of design catalogues, websites, and expert advice to design DCE questionnaires.

Design catalogues were used in a small number of studies although this has risen in recent years; 2 / 34 (6%) of studies in 1990 – 2000, 6 / 114 (5%) in 2001 – 2008, rising to 15 / 96 (16%) in 2009 – 2011. Although websites were not used to source DCE designs in 1990 – 2000, they accounted for 3 / 114 (3%) of studies in 2001 – 2008, rising slightly to 4 / 96 (4%) in 2009 – 2011. Finally, use of expert advice to underpin study design varied. It accounted for 4 / 34 (12%) of cases in 1990 – 2000, down to 4 / 114 (4%) in 2001 – 2008, before rising again to 6 / 96 (6%) in 2009 – 2011. Unfortunately, the 'Design Source' was not clearly reported in a significant proportion of studies. It accounted for 9 / 34 (26%) of studies in 1990 – 2000, rising to 42 / 114 (37%) of cases in 2001 – 2008, before falling back to 26 / 96 (27%) in 2009 – 2011.

D.4.4. Methods used to create choice sets.

In 1990 – 2000, 3 / 34 (9%) of designs involved Orthogonal arrays with single profiles (i.e. binary choices), the figure rose to 12 / 114 (11%) in 2001 – 2008, but falling to 2 / 96 (2%) in 2009 – 2011. Use of orthogonal arrays with random pairing was more common but has fallen over time; in 1990 – 2000 18 / 34 (53%) of analyses , falling to 19 / 114 (17%) in 2001 – 2008, and then falling again to 11 / 96 (11%) in 2009 – 2011.

Studies involving orthogonal arrays with pairing with a constant comparator constituted about one in five designs at first (6 / 34 (18%) in 1990 – 2010 and 23 / 114 (20%) in 2001 – 2008) before falling to 3 / 96 (3%) in 2009 – 2011 (during a period when this approach became

more open to criticism). In 1990 – 2000 no studies involved orthogonal arrays with foldover random pairing, the figure remained low, 1 / 114 (1%) in 2001 – 2008 and 1 / 96 (1%) in 2009 – 2011. In 1990 – 2000, none of the studies reviewed involved orthogonal arrays with foldover, but this figure rose to 11 / 114 (10%) in 2001 – 2008, and rose further to 20 / 96 (21%) in 2009 – 2011. Similarly, there has been a general trend towards D- efficient designs, rising from 0% in 1990 – 2000 to 14 / 114 (12%) in 2001 – 2008, and 18 / 114 (19%) in 2009 – 2011. Finally, in 1990 – 2000 4 / 34 (12%) of studies were classified as 'Other (pragmatically chosen)', falling to 2 / 114 (2%) in 2001 – 2008, before rising to 6 / 96 (6%) in 2009 – 2011. The proportion of studies which did not clearly report the methods that they used to create choice sets has risen over the last two decades, from 3 / 34 (9%) in 1990 – 2000. to 32 / 114 (28%) in 2001 – 2008, and 30 / 96 (31%) in 2009 – 2011.

Table II. Background information on DCE studies.

Item	**Category**	**Baseline: 1990 – 2000**	**Period: 2001 – 2008**	**Period: 2009 – 2011**
		N = 34	N = 114	N = 96
Design type	Full factorial	4 (12%)	0 (0%)	4 (4%)
	Fractional factorial	25 (74%)	114 (100%)	83 (86%)
	Not clearly reported	5 (15%)	0 (0%)	9 (9%)
Design plan	Main effects only	25 (74%)	100 (89%)	68 (71%)
	Main effects, 2 or more way interactions	2 (6%)	6 (5%)	11 (11%)
	Not applicable	4 (12%)	0 (0%)	1 (1%)
	Not clearly reported	3 (9%)	8 (7%)	16 (17%)
	Software package	19 (56%)	59 (52%)	44 (46%)
Design Source	SPEED	13 (38%)	22 (19%)	6 (6%)
	SPSS	2 (6%)	14 (12%)	7 (7%)
	SAS	0 (0%)	14 (12%)	18 (19%)
	SAWTOOTH	2 (6%)	5 (4%)	5 (5%)
	Other	2 (6%)	0 (0%)	7 (7%)

Table II (contd). Background information on DCE studies.

	No further details	0 (0%)	4 (4%)	2 (2%)
	Catalogue	2 (6%)	6 (5%)	15 (16%)
	Website	0 (0%)	3 (3%)	4 (4%)
	Expert	4 (12%)	4 (4%)	6 (6%)
	Not clearly reported	9 (26%)	42 (37%)	26 (27%)
Method to create choice sets	Orthogonal arrays: Single profiles (i.e. binary choices)	3 (9%)	12 (11%)	2 (2%)
	Orthogonal arrays: Random pairing	18 (53%)	19 (17%)	11 (11%)
	Orthogonal arrays: Pairing with constant comparator	6 (18%)	23 (20%)	3 (3%)
	Orthogonal arrays: Foldover – random pairing	0 (0%)	1 (1%)	1 (1%)
	Orthogonal arrays: Foldover	0 (0%)	11 (10%)	20 (21%)
	D – efficiency (SAS)	0 (0%)	14 (12%)	18 (19%)
	Other (pragmatically chosen)	4 (12%)	2 (2%)	6 (6%)
	Not clearly reported	3 (9%)	32 (28%)	30 (31%)
	Other	N / A	N / A	14 (15%)

D.5. Estimation procedures.

D.5.1. The importance of Random Utility Theory (RUT).

A good overview of the cases for using different types of econometric methods in order to analyse DCE data is provided by de-Bekker Grob et al (de Bekker-Grob, Ryan et al. 2012). The authors begin by pointing out that DCE choices need to be modelled within the framework of Random Utility Theory (RUT) (McFadden 1974) In this utility framework the latent utility of alternative i in choice set C_n (for individual n) can be decomposed into two parts $V(X_{in,}\beta)$, a systematic component which can be explained as a function of changes in attribute levels, and ε_{in} which relates to unmeasured variation in preferences, and in this framework the utility of alternative i for individual n is given by:

$$U_{in} = V(X_{in,}\beta) + \varepsilon_{in}$$

(equation 16)

RUT models as de-Bekker Grob et al (de Bekker-Grob, Ryan et al. 2012) point out, assume individual n chooses alternative i if that alternative maximises utility among all alternatives within the choice set $C_{n.}$

D.5.2. Binary Probit and Logit and Multinomial Logit (MNL).

As de-Bekker Grob et al (de Bekker-Grob, Ryan et al. 2012) point out, when DCEs present binary / forced choices (you select from one of two options) then binary probit or logit models can be used, however once more than 3 options are presented then multinomial logit (MNL)

becomes what they describe as the “workhorse for data analysis.” However as de-Bekker Grob et al (de Bekker-Grob, Ryan et al. 2012) also point out, MNL requires 3 assumptions:

(i) Independence of irrelevant alternatives (IIA)
(ii) The error terms are independent and identically distributed (IID) across observations
(iii) There is no taste heterogeneity (i.e. preferences are homogeneous across respondents).

The authors suggest that “Such assumptions may be restrictive in describing human behaviour, questioning the validity of policy conclusions.”

D.5.3. Other models which allow the assumptions of MNL to be relaxed.

Other models (see figure 2 in (de Bekker-Grob, Ryan et al. 2012)) can be used for applications which do not conform to the 3 underlying assumptions of MNL as follows:

(i) Models which relax the independence of irrelevant alternatives assumption (IIA): Such models involve more flexible error distributions and include models such as Nested Logit and multinomial Probit.

(ii) Models which relax the independent and identically distributed (IID) assumption: Such as heteroscedastic models, i.e. heteroscedastic logit which allows for flexible as opposed to a fixed variance.

(***iii) Models which allow for taste heterogeneity:*** These include Mixed Logit and Latent class models (LCM) which allows for taste variation. Mixed Logit (also known as random

parameters logit) avoids potential estimation bias from unobserved preference heterogeneity in discrete choice models by estimating a distribution of preferences across patients for each preference parameter (Hauber, Gonzalez et al. 2011). With LCM parameter heterogeneity across individuals is modelled with a discrete distribution over a set of classes, and individuals are sorted into a set of classes (Mentzakis, Ryan et al. 2011)

D.5.4. Use of Probit, Random Effects Probit, Logit, and Random Effects Logit.

As previously reported, early DCE studies i.e. 1990 – 2000 (Ryan and Gerard 2003) seemed to focus upon applying either binary choice or 'forced choice' DCEs. So, for example in 1990 – 2000, 6 / 34 studies (18%) were Probit; this figure fell to 8 / 114 (7%) in 2001 – 2008, and fell further to 2 / 96 (2%) in 2009 – 2011. Similarly, in 1990 – 2000 18 / 34 studies (53%) were Random Effects Probit, falling to 47 / 114 (41%) in 2001 – 2008, and then further to 14 / 96 (15%) in 2009 – 2011. The number of Logit studies was 1 / 34 (3%) in 1990 – 2000, rising to 13 / 114 (11%) in 2001 – 2008, and falling to 4 / 96 (4%) in 2009 – 2011. The number of Random Effects Logit studies was similarly small; 1 / 34 (3%) in 1990 – 2000, 6 / 114 (5%) in 2001 – 2008, and 9 / 96 (9%) in 2009 – 2011.

D5.5. Use of Multinomial Logit (MNL).

It would appear that the decline in the use of Logit, and Probit and Random Effects Probit reported above has been offset by an increased use of MNL studies, which have the advantage that they can cater for more than 2 response options. During 1990 – 2000, 6 / 34 (18%) used Multinomial Logit (MNL), rising to 25 / 114 (22%) during 2001 – 2008, and increasing further during 2009 – 2011 to 43 / 96 (45%).

D.5.6 Use of Nested Logit (NL).

During the period 2001 – 2008 (de Bekker-Grob, Ryan et al. 2012) a small shift towards use of Nested Logit (a technique which relaxes the IIA assumption) was observed. It involved 5 / 114 (4%) during the period 2001 – 2008, up from 0% in 1990 – 2000. For the period 2009 – 2011, the figure remained low at 3 / 96 (3%) studies.

D.5.7. Models applicable when there is preference heterogeneity.

During 1990 – 2000, only 1 / 34 studies used Mixed logit, by 2001 – 2008, 6 / 114 studies (5%) used Mixed Logit. All of the Mixed Logit studies for the period 2001 – 2008 found evidence of some preference heterogeneity. During the period 2009 -2011, there was again a clear trend towards increased use of Mixed Logit; 17 / 96 studies (18%) utilised the technique, and in all cases there was evidence to suggest at least some preference heterogeneity. Also one study (Johnson, Ozdemir et al. 2010) used a Mixed Logit Hierarchical Bayesian model (MLHB), an extension of Mixed Logit modelling. This is an interesting development because with the MLHB approach (Johnson, Ozdemir et al. 2010) Mixed Logit population parameters are first estimated to indicate the distribution of tastes across the sample. These population parameters then serve as priors in a Bayesian update using information obtained from each individual's pattern of choices, and the mean of the Bayesian posterior of a parameter can be interpreted as a classical estimator (Johnson, Ozdemir et al. 2010). Another study used what they described as a Bayesian like approach which was similar to mixed logit (Wittink, Cary et al. 2010).

During the early period (1990 – 2000) no study used the latent class model (LCM). Later studies using LCM were few, but all provided evidence of preference heterogeneity. During the period 2001 – 2008 1 / 114 studies (1%) used LCM, and it also found preference heterogeneity. During 2009 – 2011, 3 / 96 (3%) of studies used LCM, and all 3 again identified evidence of preference heterogeneity.

A few studies used 'Other' estimation procedures. In 1990 – 2000 this was 1 / 34 (3%) of studies, the figure rose slightly to 4 / 114 (4%) in 2001 – 2008 and then rose again to 13 / 96 (14%) in 2009 – 2011. Reporting of the estimation procedure used has improved over time. Overall 2 / 34 (6%) of respondents did not clearly report the estimation procedure used in 1990 – 2000, the figure fell to 4 / 114 (4%) in 2001 – 2008, and then fell again to 1 / 96 (1%) in 2009 – 2011.

Table III. Estimation procedures

Item	Category	Baseline: 1990 – 2000	Period: 2001 – 2008	Period: 2009 – 2011
		N = 34	N=114	N = 96
Estimation procedure	Probit	6 (18%)	8 (7%)	2 (2%)
	Random Effects Probit	18 (53%)	47 (41%)	14 (15%)
	Logit	1 (3%)	13 (11%)	4 (4%)
	Random Effects Logit	1 (3%)	6 (5%)	9 (9%)
	MNL	6 (18%)	25 (22%)	43 (45%)
	Nested Logit	0 (0%)	5 (4%)	3 (3%)
	Mixed Logit	1 (3%)	6 (5%)	17 (18%)
	Latent class (LCM)	0 (0%)	1 (1%)	3 (3%)
	Other	1 (3%)	4 (4%)	13 (14%)
	Not clearly reported	2 (6%)	4 (4%)	1 (1%)

D.6. Validity.

D.6.1. Validity checks.

Table IV first begins by assessing whether Validity tests are applied. Tests of external validity are particularly valuable, if they can be applied because stated preferences from DCEs can then be compared with revealed preferences. However, DCEs are often applied in order to provide information in contexts in which revealed preference information is not available. Therefore, given the lack of revealed preference information, there is often little scope to conduct tests of external validity (particularly if DCEs are applied in the context of stated funded healthcare provision). This is presumably why table IV show that in 1990 – 2000 none of the studies contained a test for external validity. The figure rose to 1 / 114 (1%) in 2001 – 2008 due to a paper by Mark and Swait (Mark and Swait 2003) which compared doctors' prescribing decisions in relation to prescriptions for alcoholism with the preferences they expressed in a DCE. Given that the preference information obtained via DCEs and revealed preference data were similar the results proved reassuring. For the most recent time period (2009 – 2011), unfortunately there are no papers which compared DCE preferences with revealed preferences (the strict definition of external validity I used).

In terms of tests for internal theoretical validity, most studies included these. Overall 22 / 34 (65%) of studies in 1990 – 2000 included such a test, with the figure falling to 64 / 114 (56%) in 2001 – 2008, and then rising again to 69 / 96 (72%) in 2009 – 2011. Such tests involve an assessment of whether coefficients appear to move in line with prior expectations, and papers reported that they were generally met.

Tests for non-satiation were less frequently reported and the frequency has fallen. For the period 1990 – 2000, 14 / 34 (44%) of studies contained such a test, the figure rose slightly to 56 / 114 (49%) in 2001 – 2008, before falling again to 14 / 96 (15%) in 2009 – 2011. The decline in the use of such tests probably reflects concerns that they tend to be passed, so that they are a relatively weak test of validity. If tests of transitivity could readily be applied using DCEs, the information yielded might be more useful. However, they cannot always be readily applied, which is presumably why over the period 1990 – 2000 only 3 / 34 (9%) of studies contained a transitivity test, in 2001 – 2008, 5 / 114 (4%) of studies such a test, and during 2009 – 2011 no studies contained a transitivity test. Tests relating to Sen's extraction and contraction properties were rarely reported. During 1990 – 2000 none of the studies contained a test relating to Sen's extraction and contraction properties, the figure rose to 2 / 114 (2%) during 2001 – 2008, before falling back to 1 / 96 (1%) in 2009 – 2011.

Use of a test for internal compensatory decision making was much more frequent. In 1990 – 2000, 12 / 34 (35%) of studies involved such a test , the figure for 2001 – 2008 was 36 / 114 (32%), but in 2009 – 2011 it fell to only 14 / 96 (15%).

D.6.2. Use of qualitative methods to enhance DCE process and results.

Table IV also contains information on the use of qualitative methods to enhance DCE process and results. In 1990 – 2000 6 / 34 (18%) of studies used qualitative methods to inform attribute selection, the figure rose to 79 / 114 (69%) in 2001 – 2008, before falling back to 36 / 96 (38%) in 2009 – 2011. A similar trend is apparent in relation to the use of qualitative methods to inform attribute level selection the figure was 6 / 34 (18%) in 1990 – 2000, it rose to 38 / 114 (33%) in 2001 – 2008, before falling back to 25 / 96 (26%) in 2009 – 2011.

The use of a pre-testing questionnaire fell consistently over the period it was 16 / 34 (47%) in 1990 – 2000, it fell to 36 / 114 (32%) in 2001 – 2008, and then to 23 / 114 (24%) in 2009 – 2011. The use of de-briefing choices to help strengthen understanding rose from 0 in 1990 – 2000 to 5 / 114 (4%) in 2001 – 2008, it then fell down to 2 / 96 (2%) during 2009 – 2011.

Table IV. Validity

Item	**Category**	**Baseline 1990-2000**	**Period: 2001 – 2008**	**Period: 2009 – 2011**
		N = 34	N = 114	N = 96
Validity tests	External	0 (0%)	1 (1%)	0 (0%)
	Internal: Theoretical	22 (65%)	64 (56%)	69 (72%)
	Internal: Non-satiation	15 (44%)	56 (49%)	14 (15%)
	Internal: Transitivity	3 (9%)	5 (4%)	0 (0%)
	Internal: Sen's expansion and contraction	0 (0%)	2 (2%)	1 (1%)
	Internal: Compensatory decision making	12 (35%)	36 (32%)	14 (15%)
Use of qualitative methods to enhance DCE process & results	Increasing face validity: Attribute selection	6 (18%)	79 (69%)	36 (38%)
	Increasing face validity: Level selection	6 (18%)	38 (33%)	25 (26%)
	Increasing face validity: Pre-testing questionnaire	16 (47%)	36 (32%)	23 (24%)
	Increasing face validity: Strengthen understanding responses – Debriefing choices	0 (0%)	5 (4%)	2 (2%)

D.7. Areas of application and outcome measures.

D. 7.1. Areas of application.

Table V presents details of the areas of application for DCE studies represented in the 3 reviews. This shows the main study objectives and topic areas covered by DCEs published over the last two decades. As indicated by de-Bekker Grob et al (de Bekker-Grob, Ryan et al. 2012), although DCEs had been originally introduced into health economics primarily in order to value patient experience (Ryan 1999) there was clear evidence that the applications of DCEs had broadened out considerably by the time de-Bekker Grob (de Bekker-Grob, Ryan et al. 2012) wrote their review covering the period 2001 - 2008. Moreover, this trend continued into 2008 – 2011.

Table V shows that in 1990 – 2000, 12 / 34 studies (35%) had a main study objective which involved valuing experience factors, in 2001 – 2008, 40 / 114 studies (again 35% of the sample) had the same main study objective, however during 2009 – 2011 the figure fell to 13 / 96 (14%). The fall in the proportion of studies primarily involving the valuation of experience factors (in 2009 – 2011) is illustrative of the trend away from the use of DCEs just in order to value patient experience.

Throughout all 3 periods, approximately one in ten studies have had the main objective of valuing health outcomes. In 1990 – 2000, this figure was 3 / 34 studies (9%), during 2001 – 2008 it was 8 / 114 (7%), and in 2009 – 2011 the figure rose a little to 11 / 96 (11%).

In contrast, the proportion of DCE studies looking at trade-offs between health outcomes and experience factors has risen steadily. In 1990 – 2000, 3 / 34 studies (9%) had this as a main study objective of, during 2001 – 2008, the figure was 38 / 118 (33%), rising further to 48 / 96 (50%) in 2009 – 2011. This reflects a shift from looking at patient experience factors in isolation (down from 35% in 2001 – 2008 to 14% in 2009 – 2011) towards looking at trade-offs between health outcomes and experience factors (up from 33% in 2001 – 2008 to 50% in 2009 - 2011).

In 1990 – 2000, no studies main study objective was looking at utility weights within a QALY framework. During 2001 – 2008, 2 / 114 (2%) of studies had this as the main objective. These 2 studies used DCEs as an alternative to Standard Gamble (SG) and time-trade off (TTO) to estimate utility weights within a QALY framework (Ryan, Netten et al. 2006); and (Burr, Kilonzo et al. 2007). More recently there has been some further work in this area. In 2009 – 2011, 2 / 96 studies (2%) had this as their main objective reflecting continued interest in this research agenda. One of these studies (Lancsar, Wildman et al. 2011) looked at deriving distributional weights for QALYs using DCEs, whilst the other (Stolk, Oppe et al. 2010) used DCEs to quantify EQ-5D health states.

A small percentage of DCEs have the objective of evaluating job choices. In 1990 – 2000, this figure was 2 / 34 studies (6%), during 2001 – 2008 it was 5 / 114 (4%), again similarly in 2009 – 2011 the figure was 4 / 96 (4%). Whilst this is not a major application of DCEs, the technique can be used to help ascertain preferences for different health related jobs, thereby helping to inform human resource policies. During 2009 – 2011 the applications included attracting nurses to rural areas in Kenya, South Africa, and Thailand (Blaauw, Erasmus et al. 2010); and also attracting health workers to rural jobs in Tanzania (Kolstad 2011). Further

studies included establishing pharmacists' and student pharmacists' preferences for patient centred care in Canada (Grindrod, Marra et al. 2010), and establishing young physicians' preferences for choice of practice establishment in Germany (Gunther, Kurstein et al. 2010).

A number of studies have used DCEs in the context of developing priority setting frameworks. In 1990 – 2000, these represented 2 / 34 studies (6%), during 2001 – 2008, 6 / 114 (5%) of studies had the same objective. The figure rose marginally to 7 / 96 (7%) in 2009 – 2011. Applications in 2009 – 2011 included establishing Dutch patients' priorities for ambulatory hospital care centres (Albada and Triemstra 2009); establishing UK patients' priority criteria for kidney transplant allocation (Clark MD, Gumber AK et al. 2009); establishing Canadian patients' and healthcare professionals' preferences for priority criteria for kidney allocation and end of life care for patients with chronic kidney disease (Davison SN, Kromm SK et al. 2010); establishing Dutch healthcare professionals' priorities for reimbursing different healthcare interventions (Koopmanschap, Stolk et al. 2010); a UK general public survey of preferences for resource allocations relating to health technology appraisal (Green and Gerard 2009); establishing UK general public priorities for resource use (Watson, Carnon et al. 2011); and establishing priority setting criteria for HIV / Aids interventions in Thailand (Youngkong, Baltussen et al. 2010).

Finally, 1 / 34 studies (3%) in 1990 – 2000 had a main study objective of looking at health professionals' preferences, 17 / 114 (15%) of studies during 2001 – 2008 had this objective, and the figure fell to 7 / 96 (7%) in 2009 – 2011

So in summary the main change in areas of application between 2001 – 2008 and 2009 – 2011 has been a shift away from using DCEs to value experience factors (down from 35% to

14%) towards using DCEs to look at trade-offs between health outcomes and experience factors (up from 33% to 50%). This suggests that DCEs are still utilized to look at experience factors, but are now more often used to see how experience factors trade-off against health outcomes. This could be taken to be a reassuring finding, in that the way in which DCEs are applied now means they increasingly provide information not just about patient experience, but also about health outcomes. There have been small increases in the proportion of studies with a primary study objective such as valuing health outcomes (7% in 2001 – 2008, rising to 11% in 2009 – 2011) or developing priority setting frameworks (5% in 2001 – 2008, rising to 7% in 2009 – 2011). Such increases have been offset by a decrease in studies which look primarily at health professionals' preferences (15% in 2001 – 2008 falling to 7% in 2009 – 2011). Moreover, the proportion of studies for some areas of application has remained largely unchanged between this review and the last; for example, looking at utility weights within a QALY framework (2% in 2001 – 2008 and in 2009 – 2011) and evaluating job choices (4% in 2001 – 2008 and in 2009 – 2011.

D.7.2. Outcome measures.

Table V also provides an analysis of the main outputs presented in papers. In the past DCEs often expressed outputs in terms of the primary outcome measure of 'Per WTP unit' or 'Per time unit'. In 1990 – 2000, 10 / 34 studies (29%) used the 'Per WTP unit' outcome measure, increasing to 44 / 114 studies (39%) in 2001 – 2008, but in 2009 – 2011 the figure was only 16 / 96 (17%). The use of per unit of time as an outcome measure has also declined over the period. During 1990 – 2000 10 / 34 (29%) of papers used this outcome measure, in 2001 – 2008 the figure was 23 / 114 (20%), and it has now fallen back further to 3 / 96 (3%) in 2009 – 2011.

The proportion of DCEs using ‘Per risk unit’ as a main outcome unit is low, and this has fluctuated a little over the period. Between 1990 – 2000, 3 / 34 studies (9%) used this as their main outcome measure, during 2001 – 2008 the figure was 2 / 114 (2%), and in 2009 – 2011 the figure was 5 / 96 (5%).

Interestingly, only a minority of studies use monetary welfare measures as the main measure and this proportion has fallen a little in proportionate terms over the period. During 1990 – 2000, 5 / 34 studies (15%) involved a money welfare measure, during 2001 – 2008 the figure was 14 / 114 (12%), and during 2009 – 2011 the figure was 10 / 96 (10%).

The use of utility scores as the main outcome measure is more common and this has fluctuated over the period. In 1990 – 2000 8 / 34 (24%) of studies had utility score as the main outcome measure. The figure was 20 / 114 (18%) during 2001 – 2008, rising to 29 / 96 (30%) during 2009 – 2011.

There is also evidence that the use of an ‘Odds ratio’ has fluctuated. In 1990 – 2000 only 1 / 34 papers (3%) used odds ratios as the main outcome measure. By 2001 – 2008 the percentage had more than quadrupled to 9 / 114 (13%), it then fell back slightly to 6 / 96 (6%) in 2009 – 2011. Likewise the use of ‘Probability scores’ has fluctuated. These were the main outcome measure for 1 / 34 studies (3%) in 1990 – 2000. The figure increased to 15 / 114 (13%) in 2001, before declining to 4 / 96 (4%) during 2009 – 2011.

Finally, table V presents information on ‘Other’ outcome measures used. For the periods 1990 – 2000, and 2001 – 2009, the review authors did not use an ‘Other’ category for the

main outcome measure used. However, for this review I categorized a substantive number of studies 23 / 96 (24%) in the 'Other' category. Sometimes, this was because a study adopted 'Other' summary outcome measures. For example, some studies used an outcome measure relating to the 'relative importance' of attributes (Bederman S, Mahomed NN et al. 2009; van Helvoort-Postulart, van der Weijden et al. 2009; Grindrod, Marra et al. 2010; Johnson, Ozdemir et al. 2010; Brown, Pashos et al. 2011; Hauber, Gonzalez et al. 2011; Mohamed, Epstein et al. 2011; Muhlbacher and Nubling 2011). Other studies used an outcome measure which was a variation upon this i.e. MRS with respect to the most important attribute (Hauber, Mohamed et al. 2009), or the importance weight of a utility scale - IWQOL – Lite(Hauber, Mohamed et al. 2010). Other outcome measures utilized included choice shares (Boonen LHHM, Schut FT et al. 2009); odds ratio of WTP (Lloyd, Nafees et al. 2011); Willingness to Accept – WTA (Mentzakis, Ryan et al. 2011) or distribution weights for QALYs (Lancsar, Wildman et al. 2011).

I also classed a number of studies under the 'Other' category because they had more than one primary outcome measure, so that they could not reasonably just be subsumed under just one of the other outcome measures. Examples of these included the use of willingness to pay and willingness to wait (Chan, Sahota et al. 2009); ranks as well as utility scores (de Bekker-Grob, Essink-Bot et al. 2009); MRS as well as WTP (Deverill, Lancsar et al. 2010); incremental WTP and incremental utility (Essers, van Helvoort-Postulart et al. 2010); WTP and predicted choice probabilities for vaccines (Poulos, Yang et al. 2011); comparison of DCE choices and explicit choices (Thrumurthy, Morris et al. 2011); comparison of time trade off scores (TTO) vs DCE scores (Ratcliffe, Brazier et al. 2009); comparison of Best-worst scaling DCE versus Standard Gamble (SG) and TTO (Ratcliffe, Couzner et al. 2011); and comparison of Visual Analogue Scale (VAS), TTO, and DCE (Stolk, Oppe et al. 2010).

So in summary the main finding emerging from the 2009 -2011 review in relation to summary outcome measures is that the use of ‘Per unit WTP’ as the main outcome measure is declining (down from 39% in 2009 – 2011 to 17% in 2009 – 2011). There has also been a small decline in the percentage of studies with a summary measure which is a monetary welfare measure (down slightly from 12% in 2009 – 2011 to 10% in 2009 – 2011). Moreover, the use of ‘Per unit of time’ as the main outcome measure has also declined, but more steeply (down from 20% in 2009 – 2011 to 3% in 2009 – 2011).

In relation to the use of other summary outcome measures there does not appear to be a consistent trend so much as evidence of fluctuation. For example, use of ‘Per risk unit’ (accounted for 9% of studies in 1990 – 2000, 2% in 2001 – 2008, and 5% in 2009 – 2011); use of ‘utility scores’ (accounted for 24% of studies in 1990 – 2000, 18% in 2001 – 2008, and 30% in 2009 – 2011); use of ‘odds ratios’ (accounted for 3% of studies in 1990 – 2000, 13% in 2001 – 2008, and 6% in 2009 – 2011); and probability scores (accounted for 3% of studies in 1990 – 2000, 13% in 2001 – 2008, and 4% in 2009 – 2011).The other trend worthy of note is that, in contrast to previous review periods when no studies were categorised as using ‘Other’ outcome measures, for 2009 – 2011 I categorized 23 / 96 (24%) in this category. This was partly driven by a substantial proportion of studies utilising a mixture of outcome measures.

Table V. Output of DCEs.

	No.	Per WTP unit	Per time unit	Per risk unit	Monetary welfare measure	Utility score	Odds ratio	Probability score	Other
Main study – objective – Baseline: 1990 – 2000									
(A) Valuing experience factors	12 (35%)	3	4		1	3	1		
(B) Valuing health outcomes	3 (9%)	1				1			
(C) Trade-offs health outcomes & experience factors	14 (41%)	4	6	3	4	1		1	
(D) Utility weights within QALY framework	0 (0%)								
(E) Job-choices	2 (6%)	2							
(F) Developing priority setting frameworks	2 (6%)					2			
(G) Health professional's preferences	1 (3%)					1			
(H) Other	0 (0%)								
Total	34 (100 %)	10 (29%)	10 (29%)	3 (9%)	5 (15%)	8 (24%)	1 (3%)	1 (3%)	0 (0%)
Main study – objective – 2001– 2008									
(A) Valuing experience factors	40 (35%)	17	12		4	6		1	
(B) Valuing health outcomes	8 (7%)		1						
(C) Trade-offs health outcomes & experience factors	38 (33%)	20	6	1	10	3	4	7	

Table V (contd). Output of DCEs.

(D) Utility weights within QALY framework	2 (2%)					2			
(E) Job-choices	5 (4%)	3						1	
(F) Developing priority setting frameworks	6 (5%)	1		1		2	2	2	
(G) Health professional's preferences	17 (15%)	2	4			7	3	4	
(H) Other	4 (4%)	1							
Total	114 (100 %)	44 (39%)	23 (20%)	2 (2%)	14 (12%)	20 (18%)	9 (8%)	15 (13%)	0 (0%)
Main study – objective – 2009– 2011									
(A) Valuing experience factors	13 (14%)	3			3	4		1	2
(B) Valuing health outcomes	11 (11%)	1		1	1	3			5
(C) Trade-offs health outcomes & experience factors	48 (50%)	11	1	2	6	11	3	2	12
(D) Utility weights within QALY framework	2 (2%)								2
(E) Job-choices	4 (4%)	1				1	1		1
(F) Developing priority setting frameworks	7 (7%)		1			4	1	1	
(G) Health professional's preferences	7 (7%)		1	1		4			1
(H) Other	4 (4%)			1		2	1		
Total		16 (17%)	3 (3%)	5 (5%)	10 (10%)	29 (30%)	6 (6%)	4 (4%)	23 (24%)

D.8. Main findings of the literature review.

Ryan and Gerard (Ryan and Gerard 2003) identified 34 studies for the 1990 – 2000, implying a mean of just over 3 papers per year. The more recent review (de Bekker-Grob, Ryan et al. 2012) identified 114 DCEs relating to the period 2001 - 2008. This was equivalent to a mean of 14 studies per year. For the period 2009 – 2011 my review sourced 96 DCE studies, implying an average of 32 per year during this period. I also found that 40 / 96 of these studies were published in 2011, which suggests there is an ongoing trend towards increasing use of DCEs in health which shows no signs of abating. There has also been a shift away from UK dominance in terms of DCE publications. During 1990 – 2000, 59% of studies came from the UK, this figure fell to 48% in 2001 - 2008, and then to 21% in 2009 – 2011. Although the UK just remains the largest source of DCE studies (thanks to Scotland being part of the UK), it came close to being overtaken by the Netherlands and the USA which accounted for 19% and 17% of DCE studies respectively during 2009 – 2011. Moreover, the range of countries producing DCE studies has widened further when comparing 2001 – 2008 publications with those in 2009 – 2011.

de-Bekker Grob et al (de Bekker-Grob, Ryan et al. 2012) noted a wide range of policy applications for which DCEs were applied during 2001 – 2008. Our findings suggest this remains the case during the period 2009 – 2011. Moreover, de Bekker Grob et al (de Bekker-Grob, Ryan et al. 2012) also noted that the valuation of patient experience continued “to be the focus of the majority of studies” and found that 35% of studies had this as the main study objective. In contrast, during 2009 – 2011 I found a marked decline to 14% in the proportion of studies that had patient experience as the main study objective. However, most DCEs continued to include attributes relating to patients’ experiences, but it was just that more of

these looked at trade-offs between health outcomes and experience factors. Such studies accounted for 33% of studies in 2008 – 2011, rising to 50% in 2009 – 2011.

When comparing my findings with those for the 2001 – 2008 period (de Bekker-Grob, Ryan et al. 2012), I identified that the proportion of studies with a main study objective relating to health outcomes had grown a little from 7% in 2001 – 2008 to 11%. Also, the proportion of studies relating to priority setting grew very slightly from 5% in 2001 – 2008 to 7% in 2009 - 2011. But, the proportion of studies mainly relating to health professionals' preferences fell from 15% to 7%. Whilst the proportion of studies with main study objectives relating to Utility weights within a QALY framework (2%), job choices (4%). and 'Other' applications did not change between 2001 – 2008 and 2009 – 2011.

The 2001 - 2008 review (de Bekker-Grob, Ryan et al. 2012) reported that "Willingness to Pay continued to be a commonly used output from DCEs" for the period 2001 – 2008. However, the present review found evidence that the proportion of studies using either 'Per WTP unit' (39% in 2001 – 2008, and 17% in 2009 – 2011) or a 'Monetary welfare measure' (12% in 2001 – 2008, and 10% in 2009 – 2011) as the primary outcome has fallen as a percentage of overall DCE use. At the same time, given that the number of DCEs per year had increased from 14 to 32 per year between the two time periods, this does not mean the average number of papers per year with either a 'Per WTP unit' or 'Monetary welfare measure' as the main outcome measure has declined. Rather, they simply account for a smaller proportion of the growing annual number of published DCEs. Of course, the use of DCEs to elicit WTP estimates is encouraged by those who posit that WTP estimates might be of use to inform the benefits side of a Cost-Benefit Analysis (McIntosh 2006). However, numerous concerns have been raised in relation to the use of DCEs to elicit WTP. These include fear that respondents

might base their valuations more upon what they perceive something might cost or should cost, rather than registering a response which relates to the maximum amount they would be willing to pay (Ratcliffe 2000); (de Bekker-Grob, Ryan et al. 2012). Indeed in a test for consistency in willingness to pay experiments (Ryan and San Miguel 2000) which posited that if commodity A is preferred to B, then individuals WTP for A should exceed B, 30% of respondents failed the consistency test. This result was thought to be partly attributable to respondents registering cost-based responses. If respondents are responding in this way, then it is a cause for concern because it means that estimates of WTP obtained from respondents may fail to accurately include consumer surplus, which estimated WTP should accurately encompass if it is to be of used for the purposes of conducting Cost-Benefit Analyses.

Other concerns have been raised about whether estimates of WTP obtained via DCEs may be sensitive to the range specified for the monetary attribute or the presence or absence of payment per se (Skjoldborg and Gyrd-Hansen 2003). In a subsequent analysis the authors found evidence that respondents might be more influenced by the presence or absence of a non-zero cost than the level of cost indicated by the monetary attribute (Gyrd-Hansen and Skjoldborg 2008)). Other evidence suggests that the way attributes are “framed” in a DCE questionnaire (i.e. number of polyps found versus number of polyps missed etc) may impact upon estimated WTP (Howard and Salkeld 2009). Also there is reason to suppose that the way monetary attributes are described (see section E1.2) may affect how changes in the monetary attribute are valued.

There are also strong reasons to suppose that there might be problems with hypothetical bias when DCEs are used to estimate WTP in contexts where healthcare is free at the point of use (see section E1.1 of this chapter). Also, WTP estimates are likely to vary across different respondent groups. When Cost-Benefit Analysis is deployed the relevant perspective to use to inform resource allocation decisions is that of the general public. However, most DCE studies in health relate to patient preferences. This means that patient DCEs which generate WTP estimates may not be of use for the purposes of Cost-Benefit Analysis.

Another issue arises because when estimating MWTP, it is commonly assumed that marginal utility of money is constant and the cost function is therefore assumed to be continuous and linear. However, there is reason to think that the cost gradient may not be continuous and linear, as is commonly supposed when estimating WTP using DCEs (Johnson, Mohamed et al. 2011). Moreover, alternative specifications for estimating cost generate different estimates of WTP and respondents may deploy heuristics which involve recoding costs into categories such as low, medium, and high, thereby undermining the validity of WTP estimates (Johnson, Mohamed et al. 2011). It therefore has been argued that ‘cheap talk’ or graphic representations of attribute cost levels in questionnaire preambles may improve response validity (Johnson, Mohamed et al. 2011). Indeed some evidence (Ozdemir, Johnson et al. 2009), has indicated that the cost function for a ‘cheap talk’ subsample of DCE respondents, appears to be linear whereas the cost function for the main sample of respondents not engaged in ‘cheap talk’ is not.

It could be that these wide ranging concerns, which raise question marks about the methodological robustness of establishing WTP using DCEs, have made people more wary of

conducting WTP analysis using DCEs both in general, and also particularly in contexts in which healthcare is free at the point of use (where estimates of WTP may be more susceptible to hypothetical bias). It is also conceivable that, as more papers have been published which raise methodological concerns about WTP analysis (also see section E1.1 of this chapter), both peer reviewers and the editorial staff of journals have become more discriminating when it comes to evaluating whether a DCE which uses WTP analysis is worthy of publication.

The authors of the 2001 – 2008 review (de Bekker-Grob, Ryan et al. 2012) suggested that their findings indicated that a "fewer number of attributes are used in current DCEs". However, my review for 2009 – 2011 uncovered a more complex pattern in relation to trends in the numbers of attributes included in studies. The percentage of studies with only 2 – 3 attributes has declined from 15% during 2001 – 2008, to 13% in 2001 – 2008, and again to 10% in 2009 – 2011. However, the percentage of studies with 4 -5 attributes has fluctuated (29% in 1990 – 2000, 44% in 2001 – 2008, and 30% in 2009 – 2011). There was also a small increase in the percentage of studies with 6 attributes (26% in both 1990 – 2000, and 2001 – 2008, but up to 29% in 2009 – 2011), and a substantial increase in the proportion of studies with 7 – 9 attributes (12% in 1990 – 2000, 13% in 2001 – 2008, but up to 23% in 2009 – 2011). The percentage of studies with 10 attributes was 6% in 1990, and has remained at 2% for both 2001 – 2008 and 2009 – 2011. At the same time, the percentage of studies with more than 10 attributes fell from 12% in 1990 – 2000 to 2% in 2001 – 2008 and has risen to 5% in 2009 – 2011. Thus, if anything the trend between 2001 – 2008, and 2009 – 2011 has been towards including more attributes rather than fewer in DCEs. Moreover there has also been an interest in developing approaches to cater for the inclusion of increased numbers of

attributes within DCE designs (van Helvoort-Postulart, Dellaert et al. 2009); Witt, Scott et al. 2009).

Bringing together my findings and those of earlier reviews, there is obviously a long established pattern of DCEs typically covering a range of possible attribute domains (both in 1990 – 2000, 2001 – 2008, and 2009 - 2011) with a substantive proportion of DCEs focusing on a monetary measure, time, risk, health status, and health care related domains. Moreover, increasing numbers of studies are related to 'other' or several domains, (9% in 1990 – 2000, 15% in 2001 – 2008, and 41% in 2009 – 2011). There has also been a marked fall in the proportion of DCEs presenting 8 or fewer DCE choices (38% in 1990 – 2000, 39% in 2001 – 2008, but down to 23% in 2009 – 2011). This trend has been offset by a recent increase in the percentage of studies presenting 9 – 16 choices (53% in 1990 – 2000, 38% in 2001 – 2008, up to 55% in 2009 – 2011) whilst the proportion of studies with more than 16 choices has not changed much since 2001 – 2008 (6% in 1990 – 2000, 18% in 2001 – 2008, and 19% in 2009 – 2011).

In terms of questionnaire administration, the shift away from self-administration of DCEs (79% in 1990 – 2000, 67% in 2001 – 2008, and 60% in 2009 – 2011) has arisen largely due to an increase in the percentage of DCEs administered via a computer (9% in 1990 – 2000, 11% in 2001 – 2008, and 26% in 2009 – 2011). Increasing respondent access to computers and e-mail means this trend is likely to continue. These developments might be particularly helpful if there is a need to represent some attributes such as risk in a pictorial form as an aid to comprehension. It may also reduce the cost of accessing respondents and administering DCEs in some cases.

The trend towards the increased use of 'efficient' DCE designs noted by de-Bekker Grob (de Bekker-Grob, Ryan et al. 2012) has continued. In 1990 – 2000, D-efficient designs were not used. During 2001 – 2008, 11% of studies used D-efficient designs, a figure which has further risen to 19% in 2009 - 2011. Although there has been an increase in the percentage of studies catering for interactions (6% in 1990 – 2000, 5% in 2001 – 2008, and 11% in 2009 – 2011), main effects designs remain the dominant form of design. However, as the authors of the 2001 – 2008 review (de Bekker-Grob, Ryan et al. 2012) have argued, future work still needs to increasingly explore the inclusion of interactions terms in the design and analysis of DCEs, and more exploration is needed to assess the impact of including prior knowledge of parameters (perhaps from DCE pilot studies) within final 'efficient' DCE designs.

Interestingly, there has been a shift away from using estimation procedures such as Probit, and Random effects Probit. This is in part due to the fact that a lower proportion of DCEs involve the presentation of binary choices. The use of Multinomial Logit (MNL) models has correspondingly increased, so that they account for 45% of studies in 2009 – 2011 (up from 18% in 1990 – 2000, and 22% in 2001 – 2009). For the first time, in 2001 – 2008, 3% of studies applied Nested Logit, during 2009 – 2011 the figure was 4%.

The interest in applying models which cater for preference heterogeneity has increased (see also section E1.3). In 1990 – 2000, Mixed Logit studies accounted for 3% of studies, 5% in 2001 – 2008, but 18% in 2009 – 2011. There were no latent class model DCE studies in 1990 – 2000, but these accounted for 1% of studies in 2001 – 2008, and 3% in 2009 – 2011. I anticipate that the trend towards the increased use of models that cater for preference heterogeneity will continue. These should be applied when preferences are not homogenous and, when they have been applied, they have usually indicated that preferences are

heterogeneous (suggesting that their use is appropriate). There have also been some notable developments in this area including the use of Mixed Logit Hierarchical Bayesian models (Johnson, Qian et al. 2010) and adoption of Bayesian like approaches similar to mixed logit (Wittink, Cary et al. 2010).

Whilst validity checks are still often applied and the use of some validity checks has increased since 2001 – 2008 (i.e. internal theoretical validity checks) there appears to have been a decline in the use of most tests during 2009 – 2011. This includes tests of non-satiation, transitivity, Sen's expansion and contraction properties, and tests of compensatory decision making. We might have anticipated a decline in the use of such validity checks, because the usefulness of the results they yielded has increasingly been called into question (Lancsar and Louviere 2006); (Miguel, Ryan et al. 2005) (Ryan and Watson 2009)). However, one finding I was unprepared for and I regard as a cause for concern, is the decline in recent years in the percentage of studies using qualitative methods to enhance DCE process and results compared with 2001 – 2008. This is apparent in terms of all 4 measures specified in the 2001 – 2008 review. These include the use of qualitative methods to inform attribute selection; use of qualitative methods to inform level selection; use of pilot pre-testing questionnaires; and the use of qualitative methods to strengthen understanding of responses (including de-briefing choices). I suspect this trend may in part be attributable to the fact that an increasing proportion of studies are emanating from countries with a limited track record of DCE research. This may mean that those conducting them are less aware of the importance of qualitative methods to enhance DCE process and results.

In conclusion the use of DCEs in healthcare continues to grow, and the range of applications is growing, in an expanding range of countries. There is increasing evidence that more

sophisticated approaches to both DCE designs and analysis of data are improving the quality of final analyses. That said, recent evidence indicating that some qualitative approaches to improving DCE design validity are being neglected is a major cause for concern.

E.1. Areas of research addressed in this thesis.

Inevitably, when writing a thesis in an area as wide ranging as the application of DCEs within health, it is essential to home in on one or more key issues. For this thesis, I decided to pursue the following four main issues (E.1.1 – E.1.4):

E.1. Calculating Willingness to Pay using a monetary attribute and Hypothetical bias.

Any divergence between participants' stated preferences and how they would respond in a real life scenario can be defined as hypothetical bias (Champ and Bishop 2001). Hypothetical bias therefore relates to the difference between actual behaviour and hypothetical statements of value (Little and Berrens 2004; Mozumder and Berrens 2007). It implies that the hypothetical nature of the preference elicitation has biased responses (Guzman and Kolstad 2007).

DCEs are a stated preference technique, and a form of contingent valuation (CV). It has been suggested that because DCEs closely mirror actual consumer purchasing situations, they may be less prone to one of the primary drawbacks of contingent valuation, namely hypothetical bias (Lusk and Schroeder 2004). DCEs may be less susceptible to various sources of bias than other forms of WTP analysis. For example 'Yea saying' (Brown, Champ et al. 1996) which involves an over readiness to agree to a given suggested WTP. This occurs when respondents face a stark 'all or nothing' choice. However, DCEs require a

hypothetical trade off between levels of attributes. As this does not involve all or nothing choices, it makes 'Yea saying' less likely (Hanley, Wright et al. 1998). Moreover, obtaining a 'warm glow' (Sen 1977); Andreoni 1990) from expressing a high WTP is more likely using open-ended or closed-ended WTP analysis, than with DCEs. This is because with DCEs respondents make multiple trades between conflicting attributes, with conflicting price differentials. This ensures respondents focus more on what they are trading for what, rather than obtaining 'warm glow.' Finally, embedding arises when respondents' valuations of a programme are insensitive to changes in the size of benefits (Kahneman and Knetsch 1992). A similar concept (scope insensitivity) refers to findings that WTP values are not sensitive to differences in the number of units (Olsen, Donaldson et al. 2004). Both these phenomena should be less of a problem with DCEs, as respondents continually relate differences in attributes (i.e. the scope of benefits) to the scope of price differentials. Therefore tests of scope are integral to DCE design (Adamowicz 1995; Hanley, Wright et al. 1998).

However, despite the fact that for the aforementioned reasons we might expect DCEs to be less susceptible to hypothetical bias than other forms of CV, concerns remain that WTP estimates generated using DCEs might still be subject to some hypothetical bias. WTP estimates may be more subject to hypothetical bias if DCEs are applied to estimate WTP for healthcare provision which is free at the point of use. Indeed respondents may interpret cost as something that can be ignored as they in reality don't directly face a real cost (Ratcliffe 2000). If this happens and some respondents ignore the monetary attribute when they complete a DCE, it would mean the utility value of currency unit will be lower, and therefore estimates of MWTP would be correspondingly higher. They may also base their monetary valuation upon what they consider something might cost (so called 'cost-based' responses)

rather than what they are willing to pay (Ratcliffe 2000). This is of concern because 'cost-based' responses may fail to incorporate the element of consumer surplus that estimates of Willingness to Pay should encompass.

Other evidence suggests (Gyrd-Hansen and Skjoldborg 2008) that respondents might be more influenced by the presence or otherwise of a non-zero cost than the actual cost implied by a monetary attribute. This may be attributable to a lexicographic aversion to the "concept of paying" (Gyrd-Hansen and Skjoldborg 2008). Logically we might expect this to be a more pronounced influence upon hypothetical choices when respondents are used to receiving something 'free at the point of use.' This is because respondents are more likely to have a lexicographic aversion to the "concept of paying" if they are asked to express their WTP for something which they are accustomed to receiving 'free at the point of use.' If in contrast they faced a real situation in which they either had to pay for something which is currently free at the point of use in order to access it, their responses might differ from those expressed in a DCE (which would indicate the presence of hypothetical bias).

Alternatively it has been suggested that respondents may deploy heuristics which involve recoding costs into categories such as low, medium, and high (Johnson, Mohamed et al. 2011) which means that the marginal utility of money cannot be assumed to be constant. This explains some empirical evidence which suggests that the cost gradient may not be continuous and linear, as is commonly supposed when estimating WTP using DCEs (Johnson, Mohamed et al. 2011). However, it is unclear that such behaviour necessarily provides evidence of hypothetical bias. It would only be indicative of hypothetical bias if respondents applied such heuristics as a result of the hypothetical nature of DCE choices.

However, if respondents would respond in a similar way in contexts in which they actually paid for something, then it would not be hypothetical bias.

It has been argued that incorporating ‘cheap talk’ or graphic representations of attribute cost levels in questionnaire preambles, may improve the validity of responses (Johnson, Mohamed et al. 2011). There is evidence that cheap talk may alter cost function responses. Indeed it has been shown (Ozdemir, Johnson et al. 2009) that the cost function for a cheap talk subsample appeared to be linear whereas the cost function for the main sample of respondents not engaged in ‘cheap talk’ was not. Those engaged in ‘cheap talk’ in effect thought through their responses in greater depth before making a final choice. Therefore their ‘cheap talk’ responses could be more akin to those they might have registered had they regularly made such choices in a non-hypothetical situation. So it could well be that when respondents face hypothetical DCE questions, because they lack experience of making choices, the trade-offs they express are less valid. However, if they are forced to engage in ‘cheap talk’ this hypothetical bias (arising because responses to a hypothetical question may be less thought through) might be avoided.

The suggestion that hypothetical bias might arise when responses are less thought through, fits with evidence that estimates of WTP may be affected by the number of choices that respondents face (Bech, Kjaer et al. 2011)). This may be indicative of the fact that respondents engage in a learning process whilst completing a DCE questionnaire such that latter responses more accurately reflect preferences than earlier ones.

There is also evidence that the way in which hypothetical questions are framed can affect WTP results (Howard and Salkeld 2009). When assessing the costs and benefits of screening for colorectal cancer, the framing of attributes was manipulated to be positive (the number of cancers or polyps found), or negative (the number of cancers of polyps missed). Moreover, there was evidence that the way hypothetical questions were framed had a statistically significant impact upon estimates of WTP.

In chapter 2 of the thesis I demonstrate that if it can be assumed that WTP can be accurately elicited using DCEs, the information obtained might be of policy making utility. I applied DCEs in order to address the issue of whether healthcare for patients with suspected Deep Vein Thrombosis (DVT) ought to be provided via hospital outpatients departments or via community based provision. I then calculated estimated marginal willingness to pay (MWTP), and also estimated Willingness to Pay using the Compensating Variation approach. I then considered the health policy implications of findings. This chapter therefore provides an example of some of the potential that WTP analysis undertaken using DCEs might have to shed light upon health policy issues (if it can be assumed it is not subject to bias, which is the heroic assumption that underpins that analysis).

In chapter 3 of the thesis I address one possible source of hypothetical bias. I apply DCEs to the UK National Health Service to female patients who currently do not have to pay for the NHS healthcare they receive. Clearly in this sort of context if respondents face a monetary attribute, a proportion of respondents may choose to ignore it because in reality they know that they would not have to pay (i.e. because of hypothetical bias). I therefore posed a question after the DCE questions to establish whether respondents factored in the monetary

attribute into their decision-making or not. I then used interaction dummy variables to establish whether WTP varied comparing respondent groups according to how they had replied to this question. This analysis was conducted largely upon a sample of respondents who had menstrual disorders. However, I also included in our sample some DCE respondents with other miscellaneous Gynaecological problems who like the Menstrual disorder patients, faced a monetary attribute entitled ‘Cost to you.’

A similar analysis is conducted in chapter 4 on a different sample of respondents, (women with miscellaneous gynaecological problems) who this time (unlike the analysis in chapter 3) face the same DCE but with 3 possible different descriptors of the monetary attribute (Cost to you; Amount Lost; or Willingness to Pay) rather than just ‘Cost to you’ (as with respondents to the DCE whose responses are analyzed in chapter 3). Given that one of these descriptors refers to ‘Cost’ it might be thought that the descriptor ‘Cost to you’ might encourage cost based responses, in a manner that the descriptor ‘Willingness to Pay’ might not. So in chapter 4, I examined whether there is a statistically significant difference in estimated Marginal Willingness to Pay (MWTP), comparing estimates of MWTP obtained from questionnaires with each of the 3 monetary descriptors.

E.2. Description of the cost attribute.

Respondents might be concerned that expressing WTP could represent some precursor to an attempt to introduce charging (Ratcliffe 2000). This is why when applying DCEs it is not uncommon to change the description of WTP attributes. In chapter 4, I therefore provide

details of a wide range of different monetary descriptors which have been used for health related DCEs. Taking care to describe a monetary attribute appropriately is important because it may improve the incentive compatibility of the DCE questionnaire (de Trenqualye 1995). However, it remains unclear how the use of different descriptors affects estimated WTP (Ratcliffe 2000), which is why we investigated this issue in the later part of the analysis described in chapter 4

Chapter 4, conducts a similar analysis to that conducted in chapter 3 but this time purely upon a sample of patients with miscellaneous Gynaecological problems (i.e. respondents who had replied to the questionnaire which was targeted at those with 'Menstrual disorders' [the main body of the patient sample used for the analysis in chapter 3] are excluded from the sample used for data analysis in chapter 4). I established using a question in a DCE questionnaire whether respondents claimed to have taken the monetary attribute into account when registering choices in the DCE, I then established how WTP varied according to respondents answer to the question about whether or not they take the monetary attribute into account.

Having addressed this initial 'hypothetical bias' issue, I then compared estimated WTP according to which descriptor for the monetary attribute respondents faced (this time in contrast to the analysis in chapter 3 respondents faced one of three possible different descriptors for the monetary attribute (Cost to you; Amount Lost; or Willingness to Pay). Of course the description of all the other attributes and levels in the DCE questionnaire was kept the same. I wanted to establish whether the description of the monetary attribute might affect estimated WTP.

The choice of 3 descriptors for the monetary attribute was informed by feedback from interviews, obtained when the DCE questionnaire was being piloted. We chose the 3 descriptors which pilot respondents seemed to feel to be the most appropriate 3 options (Cost to you; Amount Lost; or Willingness to Pay). However, I deliberately steered away from allowing comparison of descriptors for the monetary attribute which might affect who pays. This is because logically you might expect that respondents might respond differently to an attribute which affects who pays i.e. cost to taxpayers, or cost to the NHS vs. out of pocket cost to the respondent (Bryan and Dolan 2004). What is less clear however is whether changes to the description of the monetary attribute which don't affect who pays, might affect estimates of WTP. This is an important issue because when healthcare is 'free at the point of use' it has become common place to use alternative descriptors for the monetary attribute to 'Willingness to Pay' in order to improve the incentive compatibility of the DCE. However, it is unclear whether and how this might affect estimates of WTP. Moreover, as indicated in section E.1.1, it is possible that a descriptor which refers to costs such as the descriptor 'cost to you' might encourage cost based responses. If this is the case we might expect estimates of WTP obtained using the 'cost to you' descriptor to be lower than those obtained using the 'Willingness to Pay' descriptor. The analysis in chapter 4 therefore considers whether or not this is the case.

E.3. Preference heterogeneity.

In recent years new techniques such as Mixed Logit and Latent Class Models (LCM) have been increasingly used in order to analyse DCE data econometrically (see section D5.7 of this chapter), in order to assess the extent to which respondent preferences are heterogeneous and identify unobserved preference heterogeneity. Before the development of these techniques more basic techniques tended to be deployed relating to observable preference heterogeneity, in order to establish whether preferences varied between pre-defined respondent groups including running separate econometric models for different response groups, and also using interaction dummy variables to establish whether preferences for a given attribute differed in a statistically significant manner between the base sample and a sub-group of respondents which formed an interaction dummy variable group, or between alternative interaction dummy variable groups.

My review of the literature for 2009-2011 revealed that basic techniques including separate regression models for different respondent groups were used (Essers, Dirksen et al. 2010; Faggioli, Scalone et al. 2011; Guimaraes, Marra et al. 2009; Lloyd, Nafees et al. 2011; Pereira, Mulligan et al. 2011; Scalone, Mantovani et al. 2009; Thrumurthy, Morris et al. 2011; Tinelli, Ryan et al. 2009; Torbica and Fattore 2010; van Dam, Hol et al. 2010; van Til, Stiggelbout et al. 2009). Also interaction dummy variables were used (Boonen LHHM, Schut FT et al. 2009; Clark MD, Gumber AK et al. 2009; de Bekker-Grob, Essink-Bot et al. 2009; Koopmanschap, Stolk et al. 2010; Lagarde, Smith Paintain et al. 2011; Musters, de Bekker-Grob et al. 2011; Nayaradou, Berchi et al. 2010; Nieboer, Koolman et al. 2010;

Schwappach, Mulders et al. 2011; Vroomen and Zweifel 2011; Wyatt, Batley et al. 2010), and LR tests were used to test for restrictions i.e. separate groups vs. full sample (Johnson, Hauber et al. 2009) had been used. Running separate models for different sub-groups of respondents can allow you to establish whether attributes appear to be valued differently by different respondent groups. In particular you can see whether the range of statistically significant variables differs between samples. You can also establish (for all statistically significant variables) whether there appear to be differences in the point estimates for attribute coefficients between sub-groups. If you want to establish whether any differences in the valuation of individual attributes is statistically significantly different however, then use of interaction dummy variables is indicated. For example (Clark MD, Gumber AK et al. 2009) if you wanted to establish whether there are statistically significant differences in preferences between one sub-group of the sample and the rest of the sample (in this case between non-white ethnic minority patient vs. the rest of the sample; or South Asian patients vs. the rest of the sample; or female vs. male patients) you can use interaction dummy variables. . Here Y_{ij} is the binary dependent variable, from individuals i = 1…m, for observations j = 1…n_i. The number of observations n_i varies because the i individuals do not all complete every pairwise choice (a minority of respondents don't answer all choices). The term μ_i is the random effects error term (which allows for multiple responses from i respondents), and ε_{ij} is the standard Probit error term for individuals i for j observations. In equation 17, D is a dummy variable and is equal to 1 if the respondent is in the subgroup, otherwise it is equal to 0.

$$Y_{ij} = \beta_0+\beta_1 wait_{ij}+\beta_2 tissue_{ij}+\beta_3 dependent_{ij}+ \beta_4 age_{ij}+\beta_5 disease1_{ij}+\beta_6 disease2_{ij}+\beta_7 ill1_{ij}+ \beta_8 ill2_{ij}+\beta_9 D_{ij}+\beta_{10} D_{ij}\, wait_{ij}+\beta_{11} D_{ij} tissue_{ij}+\beta_{12} D_{ij}\, dependent_{ij}+\beta_{13} D_{ij}\, age_{ij}+\beta_{14} D_{ij}\, disease1_{ij}+\beta_{15} D_{ij}\, disease2_{ij}+ \beta_{16} D_{ij} ill1_{ij}+\beta_{17} D_{ij} ill2_{ij}+\mu i+ \varepsilon_{ij}$$

(Equation 17)

It follows that if any beta β_{10} to β_{17} is statistically significant (at a given significance level) then we have evidence to suggest that preferences for a given attribute differ (comparing the preferences for a defined sub-group of respondents with the rest of the sample) at a defined significance level within the sub-group. Moreover, if you also run the analysis for the entire sample (without using dummy variables), as well as a model with interaction dummy variables, it is possible to use likelihood ratio (LR) tests to establish whether the interaction dummy variables are jointly significant (Clark MD, Gumber AK et al. 2009).

Alternatively if you are running separate econometric models for different sub-groups of respondents, you could use an LR test to establish whether there is a jointly significant difference in preferences comparing defined sub-groups of respondents and a pooled dataset, whereby all the sub-groups are pooled together (Johnson, Hauber et al. 2009). Unfortunately this approach (unlike that deployed by (Clark MD, Gumber AK et al. 2009) has the limitation that you cannot establish whether preference for particular attributes vary in a statistically significant manner between sub-groups (which is why an interaction dummy variable approach might be preferred).

In addition more sophisticated econometric modelling including Mixed Logit (Blaauw, Erasmus et al. 2010; de Bekker-Grob, Hofman et al. 2010; Eberth, Watson et al. 2009; Goto, Takahashi et al. 2011; Hauber, Mohamed et al. 2009); Howard and Salkeld 2009; Johnson, Ozdemir et al. 2010; Mohamed, Epstein et al. 2011; Oteng, Marra et al. 2011; Ozdemir, Johnson et al. 2009; Potoglou, Burge et al. 2011; Regier, Friedman et al. 2009; Scalone, Watson et al. 2011; Scuffham, Whitty et al. 2010; Sweeting, Whitty et al. 2011; van Helvoort-Postulart, Dellaert et al. 2009; van Helvoort-Postulart, van der Weijden et al. 2009; Whitty,

Scuffham et al. 2011; and Wittink, Cary et al. 2010) and LCM (Grindrod, Marra et al. 2010; Mentzakis, Ryan et al. 2011; Mentzakis, Stefanowska et al. 2011) has also been applied to analyse DCE data.

Mixed logit and Latent Class Models (LCM) both cater for unobserved preference heterogeneity. A good exposition of the statistical methodology underpinning both Mixed Logit and the LCM technique is provided in the discussion paper and published paper produced by Hole (Hole 2007; Hole 2008). So readers with an interest in the underlying statistical methodology are referred to that work. The Mixed Logit model has continuously distributed coefficients, whereas a latent class model has coefficients which follow a discrete distribution (Hole 2007). With the Mixed Logit approach information on preference heterogeneity is obtained by treating the coefficients as random rather than fixed parameters (Eberth, Watson et al. 2009). Using this approach standard deviation of coefficients can be calculated. If the standard deviation for a given attribute is statistically significant then this provides evidence of preference heterogeneity for that attribute (Eberth, Watson et al. 2009).

In contrast with Latent Class Models (LCM), econometric results differ for different discrete latent classes. The optimum number of latent classes can be determined by increasing the number of latent classes, until criteria such as the Bayesian Information criterion, or Akaike Information criterion, suggest the number of classes is optimal. With LCM respondents have a defined probability of being in a particular latent class. Each latent class is characterized by coefficients (and confidence intervals around coefficients) for each of the attributes. These coefficients and confidence intervals are specific to that latent class. The crucial defining characteristic of the LCM technique is that the continuous distribution of heterogeneity is

approximated by a number of finite 'points of support,' which can be understood as sorting individuals into discrete classes (Greene 2007). However, 'which class contains any particular individual, whether known or not to that individual is unknown to the analyst' (Greene 2007).

In this thesis I experiment with applying interaction dummy variables to establish whether preferences vary between respondent groups. In chapter 3 interaction dummy variables are used to establish whether preferences and WTP differs between respondents who either claim they do or 'sometimes' consider the monetary attribute when making choices vs. those who say they don't consider the monetary attribute when making choices. In chapter 4 interaction dummy variables are again used to test the same hypothesis, and then additionally to establish how preferences might differ according to which monetary descriptor is used in a questionnaire The analysis in chapter 4 (in contrast to that in chapter 3) analysed DCE data from DCEs that used 3 different descriptors for WTP (Amount lost, Cost to you, and Willingness to Pay), not just 1 descriptor ('Cost to you' as in chapter 3).

In chapter 5 I again use interaction dummy variables to compare preferences for allocating renal transplants between different stakeholder groups including patients, healthcare professionals, carers, and donors / relatives of deceased donors and ethnic minority vs non-ethnic minority patients. In chapter 6, I explore diversity issues within the patient sample of renal transplant DCE respondents in more depth using interaction dummy variables. I assess whether preferences differ between the non-white patients vs. other patients; South Asian patients vs. other patients; and female vs. male patients.

In chapter 7, I apply both Mixed Logit and LCM to the same renal transplant DCE dataset, in order to assess whether these methods, or a more basic model (i.e. Conditional Logit with interaction dummy variables) yield the most useful results. In this chapter I consider whether just applying Mixed Logit or LCM analysis (without first developing a hypothesis to test about how preferences might differ) is sufficient to provide important information about preference heterogeneity. Results obtained using Mixed Logit and LCM are then compared with the results obtained from a Conditional Logit model with interaction dummy variables for ethnic minority patients. The interaction dummy variables are specified to test a key hypothesis that preferences for allocating kidneys differ between ethnic minority and non-ethnic minority patients. In chapter 8 I apply interaction dummy variables to compare responses between respondents who claim that they only considered others when expressing DCE choices (self reported 'altruistic' responders) vs. responders who report that their own self interest either partially or wholly influenced their preferences.

E.4. Altruism.

The DCEs that have been deployed in healthcare (which I reviewed in section D) tend to assume that respondents will primarily adopt a neo-classical self interested perspective when answering a DCE. However, much of the DCE analysis reported in this thesis (chapters 5, 6, 7, and 8) is applied in a context in which altruistic motivations may be of considerable importance. Indeed altruistic concerns for the welfare of others are widely cited as the main motivation for organ donation (Gill and Lowes 2008); (Patel, Chadha et al. 2011); (Siminoff, Mercer et al. 2007)). This is why I considered that assessing the possible impact of 'altruistic motivations' was very important for this thesis. Moreover, given the fact that organs are

usually freely donated, it would not be surprising if renal patients who responded to our DCE survey expressed preferences impacted by altruistic motivations, rather than just wanting organs to be allocated in a manner which is primarily to their personal advantage. So we anticipated that altruistic motivations would play a role in determining the preferences of some patient respondents, who answered our DCE questionnaire.

The concept of the 'caring externality' (Culyer 1976) allows for the possibility that the utility that an individual derives from healthcare, may not solely derive from treatment / healthcare obtained by a given individual. Rather instead it allows for the possibility that altruistic concerns (about the health status of others) might enter into individuals' utility functions. The existence of 'caring externalities' has also been used as an important justification for the NHS and welfare state more generally (Culyer 1980), and could be used to help explain preferences for organ allocation. The 'Caring externality' embraces a conception of altruism whereby individuals choose on the basis of their own preferences but their preferences include benevolent ones, such that utility is derived from the amount of utility enjoyed by others (Dowie 1985).

Whilst some economic analysis has suggested that individuals might not exhibit 'pure altruism' and individuals are generally self-seeking (Wildman and Hollingsworth 2009) other analyses seem to suggest that in healthcare caring externalities exist (Jacobsson, Carstensen et al. 2005). The authors (Jacobsson, Carstensen et al. 2005) suggest that caring externalities may be more pronounced if the ill health suffered by others is more severe. Indeed the authors conducted a WTP analysis (using mainstream WTP analysis not DCEs) whereby some respondents valued 'internal preferences' (i.e. health benefits to themselves),

whilst others valued 'caring externalities' in WTP terms (i.e. their WTP for health benefits for another unrelated individual). The authors found evidence that respondents generally valued 'caring externalities' positively. They also found that "Differences between caring externalities and internal preferences were large, and caring externalities were consistently smaller than internal preferences." (Jacobsson, Carstensen et al. 2005). Moreover, as respondents valued more severe states of ill health experienced by others, they found that "the increase in caring was higher regarding others than oneself." (Jacobsson, Carstensen et al. 2005). The fact that 'caring externalities' and therefore altruistic preferences might be more important in relation to states of severe ill health (like that experienced by many patients on dialysis) suggests that altruistic preferences / caring externalities may play an important role in relation to preferences for renal transplantation.

So establishing whether or not respondents are basing their DCE responses about preferences for prioritizing people for renal transplants either entirely or partially upon altruistic motivations, is pertinent in relation to this research. Therefore in chapter 8 of the thesis, preferences were analyzed to ascertain whether they differed between respondents who answered a question about the perspective they adopted differently. Respondents could indicate that they expressed DCE choices on an entirely altruistic basis i.e. they could tick a box to indicate they only considered what was best only for others. Alternatively, they could tick a box to indicate they considered what was best for them and others. Or they could tick a box to indicate they only considered what was best for them. I tried to establish whether preferences were affected by the motives that respondents claimed affected their decision making. Interaction dummy variables were therefore used to ascertain whether preferences differed between groups.

F. References.

The 96 papers reviewed (2009-2011) in section D.

Ahmed, A. and J. E. Fincham (2011). "Patients' view of retail clinics as a source of primary care: boon for nurse practitioners?" J Am Acad Nurse Pract **23**(4): 193-199.

Albada, A. and M. Triemstra (2009). "Patients' priorities for ambulatory hospital care centres. A survey and discrete choice experiment among elderly and chronically ill patients of a Dutch hospital." Health Expect **12**(1): 92-105.

Bederman S, Mahomed NN, et al. (2009). "In the Eye of the Beholder: Preferences of Patients, Family Physicians, and Surgeons for Lumbar Spinal Surgery." SPINE **35**(1): 108-115.

Benjamin, L., F. E. Cotte, et al. (2012). "Physicians' preferences for prescribing oral and intravenous anticancer drugs: A Discrete Choice Experiment." Eur J Cancer **48**(6): 912-920.

Bhatt, M., G. R. Currie, et al. (2010). "Current practice and tolerance for risk in performing procedural sedation and analgesia on children who have not met fasting guidelines: a Canadian survey using a stated preference discrete choice experiment." Acad Emerg Med **17**(11): 1207-1215.

Bijlenga, D., G. J. Bonsel, et al. (2010). "Eliciting willingness to pay in obstetrics: comparing a direct and an indirect valuation method for complex health outcomes." Health Econ.

Blaauw, D., E. Erasmus, et al. (2010). "Policy interventions that attract nurses to rural areas: a multicountry discrete choice experiment." Bull World Health Organ **88**(5): 350-356.

Bogelund, M., T. Vilsboll, et al. (2011). "Patient preferences for diabetes management among people with type 2 diabetes in Denmark - a discrete choice experiment." Curr Med Res Opin **27**(11): 2175-2183.

Boonen LHHM, Schut FT, et al. (2009). "Which preferred providers are really preferred? Effectiveness of insurers' channeling incentives on pharmacy choice." Int J Health Care Finance Econ **9**: 347-366.

Brown, D. S., E. A. Finkelstein, et al. (2009). "Estimating older adults' preferences for walking programs via conjoint analysis." Am J Prev Med **36**(3): 201-207 e204.

Brown, T. M., C. L. Pashos, et al. (2011). "The perspective of patients with haemophilia with inhibitors and their care givers: preferences for treatment characteristics." Haemophilia **17**(3): 476-482.

Bunge EM, de Bekker Grob EW, et al. (2009). "Patients' Preferences for Scolosis Brace Treatment." SPINE **35**(1): 57-63.

Chan, Y. M., D. S. Sahota, et al. (2009). "Chinese women's preferences for prenatal diagnostic procedure and their willingness to trade between procedures." Prenat Diagn **29**(13): 1270-1276.

Clark MD, Gumber AK, et al. (2009). "Prioritizing patients for renal transplantation?: Analysis of patient preferences for kidney allocation according to ethnicity and gender." J Diversity Health & Social Care **6**: 181-191.

Damen, T. H., E. W. de Bekker-Grob, et al. (2011). "Patients' preferences for breast reconstruction: a discrete choice experiment." J Plast Reconstr Aesthet Surg **64**(1): 75-83.

Darba, J., G. Restovic, et al. (2011). "Patient preferences for osteoporosis in Spain: a discrete choice experiment." Osteoporos Int **22**(6): 1947-1954.

Davison, S. N., S. K. Kromm, et al. (2010). "Patient and health professional preferences for organ allocation and procurement, end-of-life care and organization of care for patients with chronic kidney disease using a discrete choice experiment." Nephrol Dial Transplant **25**(7): 2334-2341.

de Bekker-Grob, E. W., M. L. Essink-Bot, et al. (2009). "Preferences of GPs and patients for preventive osteoporosis drug treatment: a discrete-choice experiment." Pharmacoeconomics **27**(3): 211-219.

de Bekker-Grob, E. W., R. Hofman, et al. (2010). "Girls' preferences for HPV vaccination: a discrete choice experiment." Vaccine **28**(41): 6692-6697.

Deverill, M., E. Lancsar, et al. (2010). "Antenatal care for first time mothers: a discrete choice experiment of women's views on alternative packages of care." Eur J Obstet Gynecol Reprod Biol **151**(1): 33-37.

Eberth, B., V. Watson, et al. (2009). "Does one size fit all? Investigating heterogeneity in men's preferences for benign prostatic hyperplasia treatment using mixed logit analysis." Med Decis Making **29**(6): 707-715.

Essers, B. A., C. D. Dirksen, et al. (2010). "Assessing the public's preference for surgical treatment of primary basal cell carcinoma: a discrete-choice experiment in the south of the Netherlands." Dermatol Surg **36**(12): 1950-1955.

Essers, B. A., D. van Helvoort-Postulart, et al. (2010). "Does the inclusion of a cost attribute result in different preferences for the surgical treatment of primary basal cell carcinoma?: a comparison of two discrete-choice experiments." Pharmacoeconomics **28**(6): 507-520.

Faggioli, G., L. Scalone, et al. (2011). "Preferences of patients, their family caregivers and vascular surgeons in the choice of abdominal aortic aneurysms treatment options: the PREFER study." Eur J Vasc Endovasc Surg **42**(1): 26-34.

Fegert, J. M., L. Slawik, et al. (2011). "Assessment of parents' preferences for the treatment of school-age children with ADHD: a discrete choice experiment." Expert Rev Pharmacoecon Outcomes Res **11**(3): 245-252.

Goto, R., Y. Takahashi, et al. (2009). "A cohort study to examine whether time and risk preference is related to smoking cessation success." Addiction **104**(6): 1018-1024.

Goto, R., Y. Takahashi, et al. (2011). "Changes in smokers' attitudes toward intended cessation attempts in Japan." Value Health **14**(5): 785-791.

Green, C. and K. Gerard (2009). "Exploring the social value of health-care interventions: a stated preference discrete choice experiment." Health Econ **18**(8): 951-976.

Grindrod, K. A., C. A. Marra, et al. (2010). "Pharmacists' preferences for providing patient-centered services: a discrete choice experiment to guide health policy." Ann Pharmacother **44**(10): 1554-1564.

Guimaraes, C., C. A. Marra, et al. (2009). "A valuation of patients' willingness-to-pay for insulin delivery in diabetes." Int J Technol Assess Health Care **25**(3): 359-366.

Gunther, O. H., B. Kurstein, et al. (2010). "The role of monetary and nonmonetary incentives on the choice of practice establishment: a stated preference study of young physicians in Germany." Health Serv Res **45**(1): 212-229.

Hauber, A. B., A. F. Mohamed, et al. (2009). "Treatment preferences and medication adherence of people with Type 2 diabetes using oral glucose-lowering agents." Diabet Med **26**(4): 416-424.

Hauber, A. B., A. F. Mohamed, et al. (2010). "Estimating importance weights for the IWQOL-Lite using conjoint analysis." Qual Life Res **19**(5): 701-709.

Hol, L., E. W. de Bekker-Grob, et al. (2010). "Preferences for colorectal cancer screening strategies: a discrete choice experiment." Br J Cancer **102**(6): 972-980.

Hong, S. H., J. Liu, et al. (2011). "Conjoint analysis of patient preferences on Medicare medication therapy management." J Am Pharm Assoc (2003) **51**(3): 378-387.

Howard, K. and G. Salkeld (2009). "Does attribute framing in discrete choice experiments influence willingness to pay? Results from a discrete choice experiment in screening for colorectal cancer." Value Health **12**(2): 354-363.

Johnson, F. R., A. B. Hauber, et al. (2009). "Using conjoint analysis to estimate healthy-year equivalents for acute conditions: an application to vasomotor symptoms." Value Health **12**(1): 146-152.

Johnson, F. R., S. Ozdemir, et al. (2010). "Effects of simplifying choice tasks on estimates of taste heterogeneity in stated-choice surveys." Soc Sci Med **70**(2): 183-190.

Kiiskinen, U., A. L. Suominen-Taipale, et al. (2010). "Think twice before you book? Modelling the choice of public vs private dentist in a choice experiment." Health Econ **19**(6): 670-682.

Kolstad, J. R. (2011). "How to make rural jobs more attractive to health workers. Findings from a discrete choice experiment in Tanzania." Health Econ **20**(2): 196-211.

Koopmanschap, M. A., E. A. Stolk, et al. (2010). "Dear policy maker: have you made up your mind? A discrete choice experiment among policy makers and other health professionals." Int J Technol Assess Health Care **26**(2): 198-204.

Kruijshaar, M. E., M. L. Essink-Bot, et al. (2009). "A labelled discrete choice experiment adds realism to the choices presented: preferences for surveillance tests for Barrett esophagus." BMC Med Res Methodol **9**: 31.

Lagarde, M., L. Smith Paintain, et al. (2011). "Evaluating health workers' potential resistance to new interventions: a role for discrete choice experiments." PLoS One **6**(8): e23588.

Lancsar, E., J. Wildman, et al. (2011). "Deriving distributional weights for QALYs through discrete choice experiments." J Health Econ **30**(2): 466-478.

Laver, K., Ratcliffe J, Stacey G, Lester L, Walker R, Burgess L, Crotty M (2011). "Early Rehabilitation management after Stroke: What do Stroke patients prefer?" J Rehabil Med **43**: 354 - 358.

Lloyd, A., P. Hodgkins, et al. (2011). "Methylphenidate delivery mechanisms for the treatment of children with attention deficit hyperactivity disorder: heterogeneity in parent preferences." Int J Technol Assess Health Care **27**(3): 215-223.

Lloyd, A., B. Nafees, et al. (2011). "Willingness to pay for improvements in chronic long-acting insulin therapy in individuals with type 1 or type 2 diabetes mellitus." Clin Ther **33**(9): 1258-1267.

Marti, J. (2011). "Assessing preferences for improved smoking cessation medications: a discrete choice experiment." Eur J Health Econ.

Mentzakis, E., M. Ryan, et al. (2011). "Using Discrete Choice Experiments to Value Informal Care Tasks: Exploring Preference Heterogeneity." Health Econ **20**(8): 930-944.

Mentzakis, E., P. Stefanowska, et al. (2011). "A discrete choice experiment investigating preferences for funding drugs used to treat orphan diseases: an exploratory study." Health Econ Policy Law **6**(3): 405-433.

Mohamed, A. F., J. D. Epstein, et al. (2011). "Patient and parent preferences for haemophilia A treatments." Haemophilia **17**(2): 209-214.

Muhlbacher, A. C. and M. Nubling (2011). "Analysis of physicians' perspectives versus patients' preferences: direct assessment and discrete choice experiments in the therapy of multiple myeloma." Eur J Health Econ **12**(3): 193-203.

Muhlbacher, A. C., I. Rudolph, et al. (2009). "Preferences for treatment of Attention-Deficit/Hyperactivity Disorder (ADHD): a discrete choice experiment." BMC Health Serv Res **9**: 149.

Musters, A. M., E. W. de Bekker-Grob, et al. (2011). "Women's perspectives regarding subcutaneous injections, costs and live birth rates in IVF." Hum Reprod **26**(9): 2425-2431.

Nayaradou, M., C. Berchi, et al. (2010). "Eliciting population preferences for mass colorectal cancer screening organization." Med Decis Making **30**(2): 224-233.

Nieboer, A. P., X. Koolman, et al. (2010). "Preferences for long-term care services: willingness to pay estimates derived from a discrete choice experiment." Soc Sci Med **70**(9): 1317-1325.

Oteng, B., F. Marra, et al. (2011). "Evaluating societal preferences for human papillomavirus vaccine and cervical smear test screening programme." Sex Transm Infect **87**(1): 52-57.

Ozdemir, S., F. R. Johnson, et al. (2009). "Hypothetical bias, cheap talk, and stated willingness to pay for health care." J Health Econ **28**(4): 894-901.

Pavlova, M., M. Hendrix, et al. (2009). "The choice of obstetric care by low-risk pregnant women in the Netherlands: implications for policy and management." Health Policy **93**(1): 27-34.

Pereira, C. C., M. Mulligan, et al. (2011). "Determinants of influenza vaccine purchasing decision in the US: a conjoint analysis." Vaccine **29**(7): 1443-1447.

Potoglou, D., P. Burge, et al. (2011). "Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data." Soc Sci Med **72**(10): 1717-1727.

Poulos, C., J. C. Yang, et al. (2011). "Mothers' preferences and willingness to pay for HPV vaccines in Vinh Long Province, Vietnam." Soc Sci Med **73**(2): 226-234.

Ratcliffe, J., J. Brazier, et al. (2009). "Using DCE and ranking data to estimate cardinal values for health states for deriving a preference-based single index from the sexual quality of life questionnaire." Health Econ **18**(11): 1261-1276.

Ratcliffe, J., L. Couzner, et al. (2011). "Valuing Child Health Utility 9D health states with a young adolescent sample: a feasibility study to compare best-worst scaling discrete-choice experiment, standard gamble and time trade-off methods." Appl Health Econ Health Policy **9**(1): 15-27.

Regier, D. A., J. M. Friedman, et al. (2009). "Valuing the benefit of diagnostic testing for genetic causes of idiopathic developmental disability: willingness to pay from families of affected children." Clin Genet **75**(6): 514-521.

Scalone, L., L. G. Mantovani, et al. (2009). "Patients', physicians', and pharmacists' preferences towards coagulation factor concentrates to treat haemophilia with inhibitors: results from the COHIBA Study." Haemophilia **15**(2): 473-486.

Scalone, L., V. Watson, et al. (2011). "Evaluation of patients' preferences for genital herpes treatment." Sex Transm Dis **38**(9): 802-807.

Schwappach, D. L., V. Mulders, et al. (2011). "Is less more? Patients' preferences for drug information leaflets." Pharmacoepidemiol Drug Saf **20**(9): 987-995.

Scotland, G. S., P. McNamee, et al. (2011). "Women's preferences for aspects of labor management: results from a discrete choice experiment." Birth **38**(1): 36-46.

Scuffham, P. A., J. A. Whitty, et al. (2010). "Health system choice: a pilot discrete-choice experiment eliciting the preferences of British and Australian citizens." Appl Health Econ Health Policy **8**(2): 89-97.

Skjoldborg, U. S., J. Lauridsen, et al. (2009). "Reliability of the discrete choice experiment at the input and output level in patients with rheumatoid arthritis." Value Health **12**(1): 153-158.

Stolk, E. A., M. Oppe, et al. (2010). "Discrete choice modeling for the quantification of health states: the case of the EQ-5D." Value Health **13**(8): 1005-1013.

Sweeting, K. R., J. A. Whitty, et al. (2011). "Patient preferences for treatment of achilles tendon pain: results from a discrete-choice experiment." Patient **4**(1): 45-54.

Swinburn, P., A. Lloyd, et al. (2011). "Preferences for antimuscarinic therapy for overactive bladder." BJU Int **108**(6): 868-873.

Thrumurthy, S. G., J. J. Morris, et al. (2011). "Discrete-choice preference comparison between patients and doctors for the surgical management of oesophagogastric cancer." Br J Surg **98**(8): 1124-1131; discussion 1132.

Tinelli, M., M. Ryan, et al. (2009). "Patients' preferences for an increased pharmacist role in the management of drug therapy." Int J Pharm Pract **17**(5): 275-282.

Torbica, A. and G. Fattore (2010). "Understanding the impact of economic evidence on clinical decision making: a discrete choice experiment in cardiology." Soc Sci Med **70**(10): 1536-1543.

Tsung-Tai C, K.-P. C., Heng-Chiang H, Lao-Nga M, and Mei-Shu L (2010). "Using discrete choice experiment to elcit doctors' preferences for the report card design of diabetes care in Taiwan - a pilot study." Journal of Evaluation in Clinical Practice **16**: 14 - 20.

van Dam, L., L. Hol, et al. (2010). "What determines individuals' preferences for colorectal cancer screening programmes? A discrete choice experiment." Eur J Cancer **46**(1): 150-159.

van Helvoort-Postulart, D., B. G. Dellaert, et al. (2009). "Discrete choice experiments for complex health-care decisions: does hierarchical information integration offer a solution?" Health Econ **18**(8): 903-920.

van Helvoort-Postulart, D., T. van der Weijden, et al. (2009). "Investigating the complementary value of discrete choice experiments for the evaluation of barriers and facilitators in implementation research: a questionnaire survey." Implement Sci **4**: 10.

van der Pol, M. and L. McKenzie (2010). "Costs and benefits of tele-endoscopy clinics in a remote location." J Telemed Telecare **16**(2): 89-94.

van der Pol, M., A. Shiell, et al. (2010). "Eliciting individual preferences for health care: a case study of perinatal care." Health Expect **13**(1): 4-12.

van Empel, I. W., E. A. Dancet, et al. (2011). "Physicians underestimate the importance of patient-centredness to patients: a discrete choice experiment in fertility care." Hum Reprod **26**(3): 584-593.

Van Houtven, G., F. R. Johnson, et al. (2011). "Eliciting benefit-risk preferences and probability-weighted utility using choice-format conjoint analysis." Med Decis Making **31**(3): 469-480.

van Til, J. A., A. M. Stiggelbout, et al. (2009). "The effect of information on preferences stated in a choice-based conjoint analysis." Patient Educ Couns **74**(2): 264-271.

Vroomen, J. M. and P. Zweifel (2011). "Preferences for health insurance and health status: does it matter whether you are Dutch or German?" Eur J Health Econ **12**(1): 87-95.

Waltzman, J. T., T. Scholz, et al. (2011). "What patients look for when choosing a plastic surgeon: an assessment of patient preference by conjoint analysis." Ann Plast Surg **66**(6): 643-647.

Watson, V., A. Carnon, et al. (2011). "Involving the public in priority setting: a case study using discrete choice experiments." J Public Health (Oxf).

Whitty, J. A., P. A. Scuffham, et al. (2011). "Public and decision maker stated preferences for pharmaceutical subsidy decisions: a pilot study." Appl Health Econ Health Policy **9**(2): 73-79.

Witt, J., A. Scott, et al. (2009). "Designing choice experiments with many attributes. An application to setting priorities for orthopaedic waiting lists." Health Econ **18**(6): 681-696.

Wittink, M. N., M. Cary, et al. (2010). "Towards Patient-Centered Care for Depression: Conjoint Methods to Tailor Treatment Based on Preferences." Patient **3**(3): 145-157.

Wyatt, J. C., R. P. Batley, et al. (2010). "GP preferences for information systems: conjoint analysis of speed, reliability, access and users." J Eval Clin Pract **16**(5): 911-915.

Yi, D., M. Ryan, et al. (2011). "Using discrete choice experiments to inform randomised controlled trials: an application to chronic low back pain management in primary care." Eur J Pain **15**(5): 531 e531-510.

Youngkong, S., R. Baltussen, et al. (2010). "Criteria for priority setting of HIV/AIDS interventions in Thailand: a discrete choice experiment." BMC Health Serv Res **10**: 197.

Other References.

Adamowicz, W. (1995). "Alternative Valuation Techniques: A comparison and movement towards a synthesis." in, K Willis and J.Corkindale (Eds.), Environmental Valuation: New Perspectives. Oxford: CAB International.

Amaya-Amaya M, K. Gerard, et al. (2008). "Discrete Choice Experiments in a Nutshell." Using discrete choice experiments in a nutshell Chapter 1, Using Discrete Choice Experiments to Value Health and Health care: 13 - 46.

Andreoni, J. (1990). "Impure Altruism and Donations to Public-Goods - a Theory of Warm-Glow Giving." Economic Journal **100**(401): 464-477.

Bech, M., T. Kjaer, et al. (2011). "Does the Number of Choice Sets Matter? Results from a Web Survey Applying a Discrete Choice Experiment." Health Econ **20**(3): 273-286.

Bradley, M. (1991). "User's manual for speed version 2.1." Hague Consulting Group, Hague.

Brown, D. S., E. A. Finkelstein, et al. (2009). "Estimating older adults' preferences for walking programs via conjoint analysis." Am J Prev Med **36**(3): 201-207 e204.

Bryan, S., M. Buxton, et al. (1998). "Magnetic resonance imaging for the investigation of knee injuries: An investigation of preferences." Health Economics **7**(7): 595-603.

Bryan, S. and P. Dolan (2004). "Discrete choice experiments in health economics. For better or for worse?" Eur J Health Econ **5**(3): 199-202.

Burgess, L. and D. Street (2005). "Optimal designs for choice experiments with assymmetric attributes." Journal of Statistical Planning and Inference **134**: 288-301.

Burgess, L. B. and D. Street (2003). "Optimal designs for 2k choice experiments." Communications in Statistics: theory and Methods **32**(11): 2185-2206.

Burr, J. M., M. Kilonzo, et al. (2007). "Developing a preference-based Glaucoma Utility Index using a discrete choice experiment." Optom Vis Sci **84**(8): 797-808.

Chakraborty, G., R. Ettenson, et al. (1994). "How consumers choose health insurance." J Health Care Mark **14**(1): 21-33.

Champ, P. A. and R. C. Bishop (2001). "Donation payment mechanisms and contingent valuation: An empirical study of hypothetical bias." Environmental & Resource Economics **19**(4): 383-402.

Coast, J. and S. Horrocks (2007). "Developing attributes and levels for discrete choice experiments using qualitative methods." J Health Serv Res Policy **12**(1): 25-30.

Culyer, A. J. (1976). "Need and the National Health Service Economics and Social choice." Oxford: Martin Robertson; 1980.

Culyer, A. J. (The political economy of social policy). "Oxford: Martin Robertson, 1980."

de Bekker-Grob, E. W., M. Ryan, et al. (2012). "Discrete choice experiments in health economics: a review of the literature." Health Econ **21**(2): 145-172.

de Trenqualye, P. (1995). "Incentive compatibility without compensation." Economic Letters **47**: 35-39.

Dowie, J. (1985). "The political economy of the NHS: individualist justifications of collective action." Soc Sci Med **20**(10): 1041-1048.

Gates, R., C. McDaniel, et al. (2000). "Modeling consumer health plan choice behavior to improve customer value and health plan market share." Journal of Business Research **48**(3): 247-257.

Gill, P. and L. Lowes (2008). "Gift exchange and organ donation: donor and recipient experiences of live related kidney transplantation." Int J Nurs Stud **45**(11): 1607-1617.

Greene, W. H. (2007). "Limdep v9.0 - Econometric Modelling Guide." Econometric Software, Inc **Vol. 1.**(New York).

Guimaraes, C., C. A. Marra, et al. (2010). "A discrete choice experiment evaluation of patients' preferences for different risk, benefit, and delivery attributes of insulin therapy for diabetes management." Patient Prefer Adherence **4**: 433-440.

Guzman, R. M. and C. D. Kolstad (2007). "Researching preferences, valuation and hypothetical bias." Environmental & Resource Economics **37**(465-487).

Gyrd-Hansen, D. and U. S. Skjoldborg (2008). "The price proxy in discrete choice experiments: Issues of relevance for future research." Ryan M., Gerard K., and Amaya-Amaya (eds.), Using Discrete Choice Experiments to Value Health and Health Care, 175-193.

Hahn, G. and B. E. Shapiro (1966). "A Catalog and Computer Program for the Design and Analysis of Orthogonal Symmetric and Asymmetric Fractional Factorial Experiments." General Electric Research and Development Center: Schenectady, NY, USA.

Hall, J., P. Kenny, et al. (2002). "Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination." Health Economics **11**(5): 457-465.

Hanley, N., R. E. Wright, et al. (1998). "Using choice experiments to value the environment - Design issues, current experience and future prospects." Environmental & Resource Economics **11**(3-4): 413-428.

Hauber, A. B., J. M. Gonzalez, et al. (2011). "Patient preferences for reducing toxicities of treatments for gastrointestinal stromal tumor (GIST)." Patient Prefer Adherence **5**: 307-314.

Hole, A. "Modelling heterogeneity in patients' preferences for the attributes of a general practitioner." CHE Research Paper 22. Centre for Health Economics. University of York, UK. 2007.

Hole, A. R. (2007). "Modelling heterogeneity in patients' preferences for the attributes of a general practitioner." CHE Research Paper 22. Centre for Health Economics(University of York, UK).

Hole, A. R. (2008). "Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment." J Health Econ **27**(4): 1078-1094.

Horowitz, J. (1985). "Random utility travel demand models." Studies in Regional Science and Urban Economics **13**: 141-151.

Jacobsson, F., J. Carstensen, et al. (2005). "Caring externalities in health economic evaluation: how are they related to severity of illness?" Health Policy **73**(2): 172-182.

Johnson, F. R., A. F. Mohamed, et al. (2011). "How Does Cost Matter in Health-Care Discrete-Choice Experiments?" Health Econ **20**(3): 323-330.

Johnson, N. M., G. Qian, et al. (2010). "Aflatoxin and PAH exposure biomarkers in a U.S. population with a high incidence of hepatocellular carcinoma." Sci Total Environ **408**(23): 6027-6031.

Kahneman, D. and J. L. Knetsch (1992). "Valuing Public-Goods - the Purchase of Moral Satisfaction." Journal of Environmental Economics and Management **22**(1): 57-70.

Lancaster, K. J. (1966). "New Approach to Consumer Theory." Journal of Political Economy **74**(2): 132-157.

Lancsar, E. and J. Louviere (2006). "Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences?" Health Econ **15**(8): 797-811.

Lancsar, E. and E. Savage (2004). "Deriving welfare measures from discrete choice experiments: inconsistency between current methods and random utility and welfare theory." Health Econ **13**(9): 901-907.

Little, J. and Berrens. R (2004). "Explaining disparities between actual and hypothetical stated values: further investigation using meta-analysis." Economics Bulletin **3**: 1 - 13.

Louviere, J., D. Hensher, et al. "Conjoint preference elicitation methods in the broader context of Random Utility Theory preferences elicitation methods." In: Conjoint measurement: Methods and applications. 2000. Chapter 13, 305-344. Heidelberg and New York: Springer.

Louviere, J. J., D. A. Hensher, et al. (2000). "Stated Choice Methods: Analysis and application." Cambridge University Press: Cambridge.

Lusk, J. L. and T. C. Schroeder (2004). "Are choice experiments incentive compatible? A test with quality differentiated beef steaks." American Journal of Agricultural Economics **86**(2): 467-482.

Mark, T. L. and J. Swait (2003). "Using stated preference modelling to forecast the effect of medication attributes on prescriptions of alcoholism medications." Value in Health **6**(4): 474-482.

Mas-Colell, A., M. D. Whinston, et al. "Microeconomic Theory. 1995." Oxford University Press.

Mc Fadden, D. "Computing Willingness-to-Pay in Random Utility Models.Trade theory and econometrics: ." In: Essays in honor of John S. Chipman. Studies in the Modern World Economy, 1999, Chapter 15, 253-74. Routledge.

McConnell, K. E. (1995). "Consumer Surplus from Discrete-Choice Models." Journal of Environmental Economics and Management **29**(3): 263-270.

McFadden, D. (1974). "Conditional logit analysis of qualitative choice behaviour." In Frontiers in Econometrics, Zarembka, P (ed.)(Academic Press: New York): 105-142.

McIntosh, E. (2006). "Using discrete choice experiments within a cost-benefit analysis framework: some considerations." Pharmacoeconomics **24**(9): 855-868.

Miguel, F. S., M. Ryan, et al. (2005). "'Irrational' stated preferences: a quantitative and qualitative investigation." Health Econ **14**(3): 307-322.

Mozumder, P. and R. P. Berrens (2007). "Investigating hypothetical bias: induced-valued voting mechanism with uncertainty." Applied Economics Letters **14**: 705-709.

Olsen, J. A., C. Donaldson, et al. (2004). "The insensitivity of 'willingness-to-pay' to the size of the good: New evidence for health care." Journal of Economic Psychology **25**(4): 445-460.

Patel, S. R., P. Chadha, et al. (2011). "Expanding the live kidney donor pool: ethical considerations regarding altruistic donors, paired and pooled programs." Exp Clin Transplant **9**(3): 181-186.

Payne, K. and R. Elliott (2005). "Using discrete choice experiments to value preferences for pharmacy services." International Journal of Pharmacy Practice **13**: 9-20.

Ratcliffe, J. (2000). "The use of conjoint analysis to elicit willingness-to-pay values. Proceed with caution?" Int J Technol Assess Health Care **16**(1): 270-275.

Ryan, M. (1999). "Using conjoint analysis to take account of patient preferences and go beyond health outcomes: an application to in vitro fertilisation." Soc Sci Med **48**(4): 535-546.

Ryan, M. (2004). "Deriving welfare measures in discrete choice experiments: a comment to Lancsar and Savage (1)." Health Economics **13**(9): 909-912.

Ryan, M. and K. Gerard (2003). "Using discrete choice experiments to value health care programmes: current practice and future research reflections." Appl Health Econ Health Policy **2**(1): 55-64.

Ryan, M. and J. Hughes (1997). "Using conjoint analysis to assess women' preferences for miscarriage management." Health Economics **6**: 261-273.

Ryan, M., A. Netten, et al. (2006). "Using discrete choice experiments to estimate a preference-based measure of outcome--an application to social care for older people." J Health Econ **25**(5): 927-944.

Ryan, M. and F. San Miguel (2000). "Testing for consistency in willingness to pay experiments." Journal of Economic Psychology **21**(3): 305-317.

Ryan, M. and V. Watson (2009). "Comparing welfare estimates from payment card contingent valuation and discrete choice experiments." Health Econ **18**(4): 389-401.

Ryan, M., V. Watson, et al. (2009). "Rationalising the 'Irrational': A Think Aloud Study of Discrete Choice Experiment Responses." Health Economics **18**(3): 321-336.

Sandor, Z. and M. Wedel (2001). "Designing conjoint choice experiments using managers' prior beliefs." Journal of Marketing Research **38**(4): 430-444.

Sandor, Z. and M. Wedel (2002). "Profile construction in experimental choice designs for mixed logit models." Marketing Science **21**(4): 455-475.

Sandor, Z. and M. Wedel (2005). "Heterogeneous conjoint choice designs." Journal of Marketing Research **42**(2): 210-218.

Sen, A. K. (1977). "Rational Fools - Critique of Behavioral Foundations of Economic-Theory." Philosophy & Public Affairs **6**(4): 317-344.

Siminoff, L., M. B. Mercer, et al. (2007). "The reasons families donate organs for transplantation: implications for policy and practice." J Trauma **62**(4): 969-978.

Skjoldborg, U. S. and D. Gyrd-Hansen (2003). "Conjoint analysis. The cost variable: an Achilles' heel?" Health Econ **12**(6): 479-491.

Sloane, N. (2009). "A library of orthogonal arrays." Available from http://www.research.att.com/~njas/oadir/.

Stensrud, J., E. Sylvestre, et al. (1997). "Targeting Medicare consumers. Managed care providers can make inroads by understanding preference and cost-sensitivity issues." Mark Health Serv **17**(1): 8-17.

Street, D. and L. Burgess (2004). "Optimal and near-optimal pairs for the estimation of effects in 2-level choice experiments " Journal of Statistical Planning and Inference **118**: 185-199.

Street, D. and L. Burgess (2007a). "The Construction of Optimal Stated Choice Experiments: Theory and Methods." Wiley: hoboken, New Jersey.

Street, D. and L. Burgess (2007b). "Discrete Choice Experiments [Computer Software]. University of Technology: Sydney [online]." Available from URL http://crsu.science.uts.edu.au/choice/ [Accessed 2010 July 13].

Street, D., L. Burgess, et al. (2005). "Quick and easy choice sets: constructing optimal and nearly optimal stated choice experiments." Internal Journal of Research in Marketing **22**: 459-470.

Street, D., L. Burgess, et al. (2008). "Designing discrete choice experiments for healthcare." In Using Discrete Choice Experiments to Value health and Health Care, Ryan M, Gerard K, Amaya-Amaya M (eds). Springer: Dordrecht, The Netherlands; 47-72.

Szeinbach, S. L., H. L. Mason, et al. (1990). "Variables affecting pharmacists' willingness to accept third-party prescription contracts: A conjoint analysis." Journal of Health Care Marketing **10**(3): 45-50.

Verbeek, M. "A Guide to Modern Econometrics. 2000. John Wiley and sons Ltd. ."

Wildman, J. and B. Hollingsworth (2009). "Blood donation and the nature of altruism." J Health Econ **28**(2): 492-503.

Zwerina, K., Huber J, et al. (2005). "A General method for constructing efficient choice designs. SAS Working Paper." Available from http://citeseerx.ist.psu.edu/viewdoc/download?doi = 10.1.1.31.9438&rep = rep 1 & type = pdf [Accessed 12 May 2009].

Chapter 2: Balancing patient preferences and clinical needs: Community versus hospital based care for patients with suspected DVT

1. Introduction.

In chapter 1 of the thesis, it was explained that there are a number of methodological problems associated with eliciting Willingness to Pay using DCEs. In this chapter such concerns are put on one side. This is in order that I can present an analysis which demonstrates how I used a DCE questionnaire (which contained a monetary attribute) to establish patients' preferences and willingness to pay (WTP) for different service models for suspected deep vein thrombosis (DVT).

Although the published analysis relating to this study (Clark, Gumber et al. 2009) simply presented estimates of Marginal Willingness to Pay (MWTP), in this chapter I also present information upon different service model uptake rates (using Value functions), and Willingness to Pay analysis using the Compensating Valuation formula.

The aim of this chapter therefore is to give a flavour of how Willingness to Pay estimates using DCEs could potentially be used to help inform decisions about which service model might best cater for patients' needs.

Issues concerning the reliability of WTP estimates, which have already been considered in chapter 1 of this thesis, are not explored in this chapter. Moreover, issues relating to the validity of obtaining estimates of WTP using DCEs are explored further in chapters 3, and 4, and the final PhD 'Discussion' chapter.

2. Background

Deep vein thrombosis (DVT) imposes a considerable burden on the healthcare system (Michota 2005). Its prevalence tends to increase in ageing populations, making it an increasing health problem. There is good evidence on optimum means of diagnosing DVT from systematic reviews and randomised controlled trials which compare different approaches (Goodacre, Sampson et al. 2006; Nijkeuter, Ginsberg et al. 2006). There is also a considerable body of evidence on its treatment.

In severe cases of DVT hospitalisation is necessary (Wichers, Di Nisio et al. 2005).Based upon a review of 17 studies, Douketis concluded that four criteria can be used to identify patients with DVT who may require hospitalisation (Douketis 2005). They included the presence of massive DVT, symptomatic pulmonary embolism, high risk of bleeding with anticoagulation, and other serious co-morbidities. It follows that for less serious cases of DVT or suspected DVT, outpatient or community based treatment should be sufficient. Indeed some estimates for the UK (Pout, Wimperis et al. 1999), suggest that around 40% of patients have 'non-complex' DVT which can be managed without a hospitalisation.

However, it is unclear whether diagnosis, and ongoing monitoring and anticoagulation therapy, should be provided via hospital outpatients departments or in the community. This study attempts to assess which of these two models might be most appropriate. Crucially, it also asks which characteristics or attributes of provision patients value most, and by how much. It follows therefore that an ultimate decision about the form of provision to be adopted may relate most to how well each of the models of provision can be tailored to meet patient preferences.

Evidence on patient preferences in relation to DVT service provision is sparse. There is some limited evidence (based on a sample of 41 patients) comparing patient preferences

for outpatient and inpatient care. Findings suggest that outpatient provision is preferred by a ratio of around 3:1 (Rodger, Gagne-Rodger et al. 2003). However, a limitation of this work, apart from the small sample size, is the lack of evidence on which characteristics of the two models of provision respondents like or dislike. The findings may therefore have been affected by differences in the perceived quality of these as provided, rather than necessarily the actual model of provision.

We have used discrete choice experiments (DCEs), a variant of conjoint analysis which is compatible with random utility theory (RUT), to consider the extent to which provision might fulfil patient preferences and how much patients with suspected DVT may value particular improvements in service provision. DCEs are of importance for decision making in contexts in which resources are scarce (Ryan and Gerard 2003),(Jones, Berney et al. 2004) and are recognised as a means of informing priority setting (Farrar, Ryan et al. 2000; Jan, Mooney et al. 2000; Morgan, Shackley et al. 2000; Schwappach 2003; Baltussen, Stolk et al. 2006). DCE methods are not used by the National Institute for Health and Clinical Excellence (NICE). NICE primarily uses evidence from cost–utility analyses which relate health outcomes (measured by quality adjusted life years [QALYs]) to costs. Using this approach, candidates for health expenditure can be ranked in terms of cost per QALY. This means that NICE is able to differentiate between interventions with 'acceptable' costs per QALY, and those deemed prohibitively expensive (i.e. using willingness to pay (WTP) per QALY). In contrast DCEs can assess the value of attributes for non-health benefits as well as attributes relating to health benefits (for example continuity of care, and avoiding hospital visits), as well those impacting upon health outcomes. WTP estimates can be used to inform the benefits side of a cost–benefit analysis (McIntosh 2006), and a strong case has been made (Ryan 2004) for NICE to consider using DCEs, but to date it has not probably because QALYs provide a standard measure allowing for assessment of efficiency across all those clinical areas for which cost-per QALY data is available

It has been argued that evidence on preferences is particularly important in relation to the assessment of diagnostic interventions (Sassi, McDaid et al. 2005) , and DCEs may significantly strengthen the existing evidence-base. In this chapter amongst other things we consider how much patients may value changes in the speed of DVT diagnosis. Quicker imaging may reduce the amount of unnecessary anticoagulation in patients who are found not to have a DVT. Prompt imaging also reduces patient anxiety. Patients may also place a value upon other aspects e.g. access, continuity of care, and fewer hospital visits, that will improve the quality of their healthcare experience. There is little research evidence on patients' valuations of such factors, or their willingness to trade off factors affecting clinical outcomes i.e. speed of diagnosis, for improvements in service provision. Crucially most clinical papers report on efficacy, based on studies in centres of clinical excellence. However, local factors e.g. how promptly a diagnosis is made, will influence actual service effectiveness. Through the application of DCEs, we are able to establish the extent to which patients would be willing to pay to avoid delays in diagnosis, and for other defined improvements in service delivery. The analysis addresses these issues.

The diagnostic pathway was not the focus of the research. Our focus was whether provision for suspected DVT ought to be provided via hospital outpatients departments or in the community. If it was via the community, the primary care trust (PCT) could negotiate 'fast-tracked' diagnosis for suspected DVT. Therefore an attribute was included to establish how much respondents valued this. The diagnostic pathway faced by respondents varied across the three Leicester hospitals. In all cases patients had an initial clinical examination, followed by D-dimer, and then either venography, or ultrasound. The latter primarily determined whether anticoagulation therapy should continue (patients would generally be put on anticoagulation if a DVT was suspected, until such a diagnosis could be largely discounted). In the pre-amble to the questionnaire, we told respondents that a decision on whether anticoagulation ought to continue would be taken using

ultrasound or venogram diagnosis. Speed of imaging diagnosis was therefore presented in terms of how quickly you had either your venogram or ultrasound. We said that about 66% of patients with suspected DVT do not have one (the figure for Leicester), and that unnecessary anticoagulation therapy could end, after an ultrasound or venogram diagnosis was established. This was a simplification because different diagnostic techniques are associated with different sensitivity and specificity. However, eliciting valuations about differing diagnostic accuracy would exceed the cognitive capabilities of many respondents so we could not probe this.

3. Materials and methods

3.1. Discrete choice experiments (DCEs)

As the literature review in section D of Chapter 1 showed, DCEs have become increasingly used within health economics. They allow patients' preferences to be elicited. Also using a series of suitably constructed scenarios it is possible to obtain information on preferences. It is also possible to conduct analysis of Marginal Willingness to Pay, Uptake rates, and calculate compensating variation (CV), a monetary measure of welfare change, if a monetary attribute is included in the DCE questionnaire. Patients value attributes and attributes define the important characteristics (Lancaster 1971) of healthcare provision which may change. By including an attribute for 'loss of income' it is also possible to estimate how much respondents might be willing to pay for defined changes in attributes (McIntosh, Donaldson et al. 1999). The key stages involved in conducting a DCE are as follows.

3.2. Identifying choice dimensions

It is important to select attributes that might be affected by the choice of providing a DVT service in the community or in a hospital outpatient department. To select appropriate attributes we conducted literature searches using MEDLINE to obtain relevant literature. We also interviewed clinicians to identify key attributes, especially those delineating the two models. Our final selection of choice dimensions included:

(a) number of patient visits to hospital;

(b) hours of available DVT nursing provision;

(c) speed of imaging diagnosis;

(d) continuity of nursing staff provision;

(e) income loss (respondents were asked to assume that "DVT provision may or may not involve you experiencing some loss of income, perhaps because of some time spent away from work")

Table 1: Details of attributes and levels.

Attribute	Lower limit level	Upper limit level	Intermediate levels	Reason for limits
Number of visits to hospital	0 visits	12 visits	4 and 8 visits	The minimum number of visits with community provision is 0. The upper limit of 12 (the limit commonly experienced by patients for the number of hospital visits).
Hours of available DVT nursing provision	5 hours	14 hours 30 minutes	9 hours and 30 minutes	5 hours was the minimum at one hospital, and 14 hours 30 minutes was the maximum at one of the hospitals
Speed of imaging diagnosis	24 hours	5 days	3 days	24 hours is usually the minimum wait. But 5 days is the maximum wait that patients experience.
Continuity of nursing staff provision	Much continuity	Some continuity	Lack of continuity	This was a discrete variable and details of the differences were explained[1]
Income loss	£0	£250	£50 and £150	This range seemed to perform well in the pilot exercise.

[1] Respondents were told that much continuity involved having the same nurse for all INR (blood) testing and counseling; some continuity was having mainly the same nurse throughout the INR testing and counseling; lack of continuity was having different nurses throughout for INR testing and counseling.

This was the first DCE I conducted (under the oversight of Emma McIntosh, University of Oxford), and we did not use mainstream Willingness to Pay analysis first in order to establish a suitable range for the monetary attribute. Instead we just tried to set what we both considered to be a plausible range for the monetary attribute.

When we discovered that the monetary attribute was highly significant in an initial pilot analysis of the questionnaire (using the range for monetary loss specified above) we concluded that the range for the monetary attribute was reasonable. If, however, I were to repeat this research again I would use mainstream Willingness to Pay analysis first, in order to ensure that the range specified for the monetary attribute was acceptable to most respondents. Community based patients may need to attend hospital only once (for imaging) and in rare cases, in which clinical diagnosis in the home is able to exclude the possibility of a DVT, may not need to attend at all. However, hospital based patients will need to attend more often. The hours of available DVT nursing provision are affected by hours prevailing in hospital DVT clinics or times when community DVT nurses are available. Speed of imaging typically varies across hospitals. Thus it will be affected by which hospital patients attend for outpatient care, or which hospital they are referred to for imaging as community based patients.

Continuity of hospital based or community based nursing provision is affected by the degree to which either form of provision can provide the same nursing staff. The extent to which patients value monetary loss may also vary across the two models.

3.3. Selection of levels for attributes

Final levels were selected for attributes to reflect rational upper and lower limits, and also sensible intermediate levels. The values selected both did not appear inappropriate and were actionable (see Table 1).

3.4. Questionnaire design: creating a DCE questionnaire

The number of scenarios that respondents need to value using DCEs is related to the number of attributes and levels assigned to each attribute (Ryan and Hughes 1997). Here the total was equivalent to 432. Thus a fractional factorial design was used (i.e. one which assumes that by obtaining information about people's responses to limited numbers of choices, the value placed on other choices can be inferred). These were generated by the computer package SPEED. When pairing off choices, it is also good practice to minimize what is known as overlap (trying to avoid pairing attributes in such a way that the levels for attributes are the same for a given attribute for both choice A and B) (Huber and Zwerina 1996) and. Emma McIntosh and I pragmatically tried to pair attributes off in such a way as to avoid overlap as much as possible. It is also good practice to avoid excesses of level imbalance (i.e. the frequency of some attribute levels appearing more often in the choices presented to respondents than other attributes) and under the direction of Dr Mcintosh, I therefore allowed some of the choices sets generated by the SPEED design template to appear more than once (to offset level imbalance in the original SPEED design), which is why we ended up with 15 pairwise choices. Reducing overlap and level imbalance should help to improve the statistical efficiency of findings (i.e. the ability of the DCE to generate information about preferences from as little response information as possible). If we could have specified 4 levels for each attribute we would have been able to generate an orthogonal design template using SPEED which did not exhibit level imbalance. However, although our pilot work indicated having 4 levels for most attributes seemed appropriate,

when it came to continuity of care, the 3 levels for continuity of care that we presented seemed more appropriate to respondents. Given we knew we would have some imbalance, we also decided to set the 3 levels for hours of available nursing provision in line with what was available via the 3 different sources of existing provision in Leicester.

Scenarios can be presented in the form of pairwise choices. We used a main effects design, thereby excluding interactions between preferences for attributes. A fixed comparator was not used; instead we had 15 pairwise choices. Moreover, we conducted the recommended “simple checks” of the correlation matrix of attribute levels and frequency checks, in order to ensure the absence of collinearity (Ryan and Gerard 2003). Following advice on best practice we also included a Consistency / dominance test (Ryan and Gerard 2003) in which respondents faced an option which for which logically we would always expect them to prefer one option to the other (i.e. if they behaved in a utility maximising manner).

3.5 Questionnaire piloting and refinement

A pilot questionnaire was administered on 30 patients. Responses were evaluated, and all but one attribute was found to be significant at the 5% level, with the other significant at the 6% level. This was a most reassuring result implying that only minor changes to the presentation had to be made. Therefore the questionnaire still had 15 pairwise choices, and the same attributes and levels.

3.6. Obtaining responses from patients

All respondents had, or previously had, a suspected DVT and attended one of the three Leicester hospitals. The questionnaire was mainly given out in hospital waiting rooms to patients with suspected DVT. However in order to reach some former diagnosed patients

(rather than just suspected DVT patients) a small sub-sample was distributed by post to patients who had previously presented with a suspected DVT in Leicester. We obtained 256/394 adequately completed responses (a usable response rate of ca. 65% overall).

3.7. Model applied

Results were generated using econometric analysis. We used Random Effects Probit. The Random effects modelling corrects the error term in the regression analysis to allow for multiple responses from respondents. Variables are defined in Table 2. The model can be expressed as:

$$Y = \alpha_0 + \alpha_1 Difcont1 + \alpha_2 Difcont2 + \alpha_3 Difvisit + \alpha_4 Difhours + \alpha_5 Difspeed + \alpha_6 Difinc + \mu + \xi$$

Here Y is the binary dependent variable, μ is the random effects error term (which allows for multiple responses from each respondent) and ξ is the other Probit error term. The constant term (α_0) if significant might potentially indicate the present of some variables omitted from the DCE design which influence choice.

Table 2. Description of variables

Attribute name	Description
Difcont1	A difference between much continuity and some continuity in DVT nursing
Difcont2	A difference between some continuity and lack of continuity in DVT nursing
Difvisit	The number of times people would have to attend hospital – 1 extra visit
Difhours	Hours during which routine dedicated DVT provision is available – 1 extra day
Difspeed	Speed of imaging diagnosis – 1 extra day's wait.
Difinc	Income loss - £1 of income loss

3.8. Calculating willingness to pay (WTP)

To estimate WTP, coefficients from the regression model are used. They indicate utility associated with changes in attributes. The coefficient upon income loss (α_6) indicates the utility of losing £1. To establish WTP for a change in one of the non-monetary attributes (associated with α_1, α_2, α_3, α_4, α_5), you need to take the coefficient you want to express in WTP terms, and divide it by α_6 (we divide by $-\alpha_6$, because what is relevant is the utility value of a pound, not a pound lost, which is what the coefficient indicates). By dividing each of the non-monetary coefficients by $-\alpha_6$, we normalize the value of the coefficients (α_1, α_2, α_3, α_4, α_5) in WTP terms[2]. It is also then possible to put 95% confidence intervals around WTP estimates using the Delta method (Wooldridge 2002), which can be executed using the non-linear confidence interval command in Stata v. 9.2. This approach to calculating WTP is applicable for 'state of the world models' whereby one alternative will be selected with certainty (Ryan 2004).

4. Results

4.1. Profile of respondents.

4.1.1. Transport and distance to hospital.

234 / 256 respondents indicated distance travelled, averaging 7.5miles. Overall 19% were within 3.5 miles; 48% within 5miles; 78% within 10 miles; and 21% travelled more than 10 miles.

[2] Note the coefficients in Table 3 follow a convention of rounding coefficients up or down (as appropriate) to 3 decimal places. However, for the purposes of deriving the actual value of willingness to pay values, we used the actual coefficients generated by the regression analysis package STATA. These are expressed to 7 decimal places, and by using these more precise point estimates, we can establish point estimates of willingness to pay for each of the non-monetary attributes, which are rounded up or down (as appropriate) to the nearest 1 pence

4.1.2. Respondents' gender and age.

241/256 respondents indicated their gender, 116/241 (ca. 48%) were male and 125/241 (ca. 52%) female. Overall 227/256 respondents indicated their age, and the average age was 58.5 years.

4.1.3. Respondents' ethnicity.

Overall 246/256 respondents indicated their ethnicity, 232/246 described themselves as white (ca. 94%); 5 as 'white and other' (ca. 2%); 3 as Black Caribbean (ca. 1%); 5 as Indian (ca. 2%); and 1 said 'Other' (ca. 0.5%).

4.1.4. Respondents' household income.

Overall 238/256 respondents indicated their gross household income band, 43/238 (ca. 18%) reported it was less than £5000; 69/238 (ca. 29%) indicated an income of £5000–11,999; 50/238 (ca. 21%) an income of £12,000–19,999; 37/238 (16%) an income of £20,000–29,999; 26/238 (ca. 11%) an income of £30,000–44,999; 7/238 (ca. 3%) an income of £45,000–59,999; and 6/238 (ca. 3%) indicated an income of £60,000+.

4.1.5. Consistency checks.

The aforementioned consistency check indicated that ca. 13% of respondents appeared to be ‘inconsistent’ responders. Although the level of ‘inconsistency’ was high, the reported data analysis was conducted upon the full data set.

4.2. Econometric results.

The attributes are described in Table 2, and the results are presented in Table 3. All the coefficients on attributes are significant at the 5% significance level (and 5 out of 6 are significant at the 1% level). Coefficients are also of the expected sign. The coefficient on having ‘much continuity’ of DVT nursing care (same nurse for INR (blood testing) and counselling throughout) rather than just ‘some continuity’ (mainly same nurse throughout INR testing and counselling) is positive. Likewise, the coefficient on ‘some’ rather than a ‘lack of continuity’ (different nurses throughout for INR testing and counselling) is also positive. The Marginal Willingness to Pay (MWTP) findings imply a positive utility premium associated with greater continuity of care of £179.32 for having much not some continuity, and £56.88 for having some not a lack of continuity. The coefficient on an extra visit to hospital is negative, suggesting utility is adversely affected by an extra visit, and respondents would pay £17.12 to avoid this. The coefficient on an extra hour of dedicated nurse provision is positive, respondents value increased availability of provision at £4.82 per hour. The coefficient on speed of imaging diagnosis is negative (having to wait an extra day for an imaging diagnosis negatively impacts on utility) and respondents would pay £115.73 to avoid this. Finally, as expected the coefficient on income loss is negative implying income loss causes disutility.

It is difficult to definitively interpret the significant constant term in the regression analysis. The fact it is significant could mean that choices are influenced by other factors apart from

attributes specified. Another possibility is that this DCE design caters for main effects but not interaction effects. If in reality there are interaction effects between particular attributes which have not been catered for in the design, the constant term could be picking this up. If I was conducting this analysis again (especially if the constant proved significant in the pilot analysis) I would use qualitative analysis to explore whether the valuation of certain attributes and their levels might interact with the valuation of others. Interaction effects, which imply that a respondent's valuation of one attribute is impacted by the levels of another attribute, may become apparent in 'think-aloud' exercises during piloting. If such effects became apparent, I would cater for interaction effects within the design (probably using a D-efficient design template in SAS).

Table 3. Regression results.

Attribute	**Coefficient**	**Confidence interval (CI) [2]**	**Marginal Willingness to Pay (MWTP)**	**CI around Willingness To Pay[3]**
Dif_cont1	.400**	(.311 / .488)	£179.32	(£123.34 / £235.31)
Dif_cont2	.127**	(.058 / .195)	£56.88	(£25.68 / £88.08)
Difvisit	-.038**	(-.031 / -.046)	-£17.12	(-£13.17 / -£21.07)
Difhours	.011*	(.002 /.020)	£4.82	(£0.65 / £8.99)
Difspeed	-.258**	(-.239 / .277)	-£115.73	(-£96.53 / -£134.93)
Difinc	-.002**	(-.002 /-.003)		
Constant	.082**	(.021 / .142)		
Mc Faddens R^2:	0.263		% of actual values predicted by the model:	50.79%

* denotes significant at the 5% level; ** denotes significant at the 1% level.
[2] 95% Confidence interval for coefficient; [3] 95% Confidence interval, using the Delta method.

4.3. Interpretation of results using Marginal Willingness to Pay (MWTP) to estimate Willingness to Pay.

If we can assume a 'state of the world' model then options can be compared using estimates of MWTP. If the alternative that the NHS provided with certainty was Model A (a home community based service requiring one visit to hospital; which was available for 9 h 30min a day; providing a diagnosis in 24 h; and providing much continuity of nursing provision) then these estimates of willingness to pay suggest that patients would be willing to pay £264.87. Alternatively, if the NHS provision available with certainty was Model B (a hospital outpatient service involving eight visits to hospital; available for 5 h; providing a diagnosis in 24 h; and providing much continuity of nursing provision), it would be valued at £123.34. Thus, patients value a community based service more.

However, if the choice was between Model B (described above) vs. Model C (a home community based service requiring one visit to hospital, available for 14 h 30min; providing a diagnosis in 24 h; but providing a lack of continuity of care). Model B would be valued at £123.34, and Model C would be valued at an average of £52.77.

Therefore these WTP results suggest that if the only 2 options available were Model A (home community based service) vs. Model B (hospital outpatient service), then community based provision (Model A) would be preferred to hospital outpatient based provision (Model B), because it is associated with a higher WTP of £264.87 rather than £123.34. However if Model A was not available and the choice was instead between Model B and C, then the hospital based service (Model B) would be preferred to the home community based service (Model C), because its associated with a WTP of £123.34, rather than £52.77 Therefore estimates of WTP can be seen to be related to how both community or hospital based provision performs in terms of the characteristics of provision

it is associated with, rather than necessarily to the community/hospital based provision distinction per se.

4.4. Value functions and uptake rates.

If we are trying to assess the relative desirability of different models of DVT provision, one way of doing this is to calculate the uptake rates for different models of provision (Ryan, Watson et al. 2008)

If we assume initially there are 2 options:

- Option 1 (Model D): A community based option which involves 1 visit to hospital; 9.5 hours of service availability; 24 hour diagnosis; and much continuity of care; income loss = £50

- Option 2: (Model E) A community based option which involves 1 visit to hospital, 14.5 hours of service delivery, 24 hour diagnosis, and some continuity of care; income loss = £50

Then the following value functions apply:

$V_{\text{Option 1}} = (-0.038 \times 1) + (0.011 \times 9.5) + (-.258 \times 1) + (0.400 + 0.127) + (-0.002 \times 50) = 0.2355$

$V_{\text{Option 2}} = (-0.038 \times 1) + (0.011 \times 14.5) + (-.258 \times 1) + (0.400) + (-0.002 \times 50) = 0.1635$

$V_{\text{No provision}} = 0$

Having calculated the value functions the probability of uptake for the 3 options can be calculated:

Pr (Option 1: Model D) = $e^{0.2355} / e^{0.1635} + e^{0.2355} + e^{0} = 0.367551582$

Pr (Option 2: Model E) = $e^{0.1635} / e^{0.1635} + e^{0.2355} + e^{0} = 0.342018103$

Pr (No provision) = $e^{0} / e^{0.1635} + e^{0.2355} + e^{0} = 0.290430315$

An additional option (Option 3) can then be introduced:

- Option 3 (Model F): A hospital outpatient based option which involves 4 visits to hospital, 14.5 hours of available service provision, 24 hours diagnosis, and much continuity of care; income loss = £0

The indirect utility of this option can then be calculated:

$V_{Option\ 3}$ = (-0.038 x 4) + (0.011 x 14.5) + (-.258 x 1) + (0.400 + 0.127) + (-0.002 x 0) = 0.2765

With the 3 options plus the no provision option then the probability of uptake changes to the following:

Pr (Option 1:Model D) = $e^{0.2355} / e^{0.1635} + e^{0.2355} + e^{0.2765} + e^{0} = 0.265776586$

Pr (Option 2:Model E) = $e^{0.1635} / e^{0.1635} + e^{0.2355} + e^{0.2765} + e^{0} = 0.247313325$

Pr (Option 3: Model F) = $e^{0.2765} / e^{0.1635} + e^{0.2355} + e^{0.2765} + e^{0} = 0.276899896$

Pr (No provision) = $e^0 / e^{0.1635} + e^{0.2355} + e^{0.2765} + e^0 = 0.21001019$

Therefore the introduction of the third option (Model F: hospital based) has altered uptake rates, and the third option (Model F) a hospital based option seems to be preferred by a higher proportion of respondents to the 2 other community based options (Models D, and E) probably because in contrast to the other options it does not entail an 'income loss'). If however we assume that the price of Model F increased (so there was an income loss of £50 not £0) then the value function for that option with a price increase would become:

$V_{Option\ 3(with\ price\ increase)} = (-0.038 \times 4) + (0.011 \times 14.5) + (-.258 \times 1) + (0.400 + 0.127) + (-0.002 \times 50) = 0.1765$

Consequently the uptake rates for the 3 options would change and would become:

Pr (Option 1:Model D) = $e^{0.2355} / e^{0.1635} + e^{0.2355} + e^{0.1765} + e^0 = 0.272969$

Pr (Option 2:Model E) = $e^{0.1635} / e^{0.1635} + e^{0.2355} + e^{0.1765} + e^0 = 0.254007$

Pr (Option 3: Model F with price increase) = $e^{0.1765} / e^{0.1635} + e^{0.2355} + e^{0.1765} + e^0 = 0.25733$

Pr (No provision) = $e^0 / e^{0.1635} + e^{0.2355} + e^{0.1765} + e^0 = 0.215694$

Therefore option 3 in the event of a price increase (income loss increases from £0 to £50) is no longer the most popular option, because option 1 is now associated with a higher uptake rate. So a community based option (option D) is then the preferred option rather than the hospital based option (option F).

Therefore in the same way in which the WTP analysis using MWTP figures indicated that the choice between community and outpatient based hospital provision was dependent upon the attribute levels of community or outpatient based hospital based provision, analysis using Value functions together with uptake rates can result in the same conclusion being reached.

Using 'uptake rates' presented in this way also has the advantage that the impact of changes in the availability of different service models, or changes in the price of service models (as proxied here by the income loss attribute) can be established upon respondents relative valuation of different models of provision. Analysis of this kind would be particularly useful to health policy makers if they were able to offer service users different models of provision, and they wanted to establish how much demand there might be for each of the different models of provision, before making them available to patients.

4.5. Compensating Variation (CV).

For the purposes of WTP analysis, if we want or need to move away from the assumption of a 'state of the world' model and assume that decision makers can choose from more than 2 alternatives we can assess welfare changes in terms of 'compensating variation' (CV).

The formula for CV is:

$$CV = 1 / -\beta_{Cost}[\ln\sum \exp(^{V1}_{j}) - \ln \sum \exp(^{V0}_{j})]$$

You can also calculate the welfare associated with the introduction of Option 3 (prior to its price increase) as follows:

$CV = 1 / 0.002 [\ln(e^{0.2355} + e^{0.1635} + e^{0.2765} + e^{0}) - \ln (e^{0.2355} + e^{0.1635} + e^{0})$

$= (500) [\ln (4.761673689) - \ln (3.443166736)]$

$= (500) [1.560599264 - 1.236391611]$

$= £162.1038265$

$= £162.10$

Therefore the introduction of Option 3 can clearly be seen to have increased welfare (i.e. it is associated with a CV of £162.10).

If in contrast option 3 was introduced at a higher price (£50 income loss not £0) then the welfare gain associated with its introduction would be lower:

$CV = 1 / 0.002 [\ln(e^{0.2355} + e^{0.1635} + e^{0.1765} + e^{0}) - \ln (e^{0.2355} + e^{0.1635} + e^{0})$

$= (500) [\ln (4.636201163) - \ln (3.443166736)]$

$= (500) [1.533895316 - 1.236391611]$

$= £148.7518525$

$= £148.75$

Using the CV method if you assumed that you started with 3 options, and wanted to identify the welfare loss associated with an increase in 'income loss' from £0 to £50 it could be calculated as follows:

$CV = 1 / 0.002 [\ln(e^{0.2355} + e^{0.1635} + e^{0.1765} + e^{0}) - \ln (e^{0.2355} + e^{0.1635} + e^{0.2765} + e^{0})$

$= (500) [\ln (4.636201163) - \ln (4.761673689)]$

$= (500) [1.533895316 - 1.560599264]$

$= -£13.35$

This result unsurprisingly is equivalent to the difference between £162.10 and £148.75 (£162.10 and £148.75= £13.35). Note the average respondents welfare loss is less than the £50 rise in the cost of the option 3 because not everyone bears the price increase (just those who still demand option 3) and also because as the price of option 3 increases, people substitute away from that option.

Crucially the information from the uptake rates (like the WTP analysis using MWTP figures) suggests that whether community or hospital based provision is the most preferred option is contingent upon the attribute levels of community or hospital based provision. Moreover, the information presented here from the CV welfare measure does not overturn this conclusion.

5. Discussion

All the analyses presented (including MWTP, Value functions and associated uptake rates, and Compensating Variation) serve to illustrate that patients' valuation of community or hospital based DVT function is inextricably related to the extent to which the different types of provision deliver a bundle of DVT service attributes that patients value. Our findings cannot simply be used to support the view that either community or hospital based DVT provision is inevitably better. This means that our findings need to be considered alongside a range of clinical considerations when deciding what type of DVT provision ought to be delivered.

Recent advancements in drug treatment both for prophylaxis and treatment of DVT (Ageno and Turpie 2005; Cohen, Hirst et al. 2005; Gutt, Oniu et al. 2005; Petersen 2005; Stannard, Lopez-Ben et al. 2006) may serve to increase the trend away from inpatient based treatment of DVT / suspected DVT. This makes consideration of whether care should be provided in hospital outpatient departments or the community an important

issue. In terms of heparin treatment, the findings of a Cochrane review of evidence in this area concluded that home treatment is no more susceptible to complications than hospital treatment. This led the reviewers to tentatively conclude that the limited evidence available indicates that home management is cost effective and (they suggested) likely to be preferred by patients (Schraibman, Milne et al. 2007).

Evidence from a well-designed Canadian economic evaluation (O'Brien, Levine et al. 1999) also supports the view that home based treatment with low molecular weight heparin is less expensive than inpatient treatment. Moreover, clinical outcomes and patient quality of life were not compromised. A more recent Spanish study (Montes, Gonzalez et al. 2005) comparing community with inpatient provision reaches the bolder conclusion that even for patients with "serious conditions" DVT management can be safely undertaken in the home and on a cost-saving basis via a home care unit.

A variety of approaches (in addition to the discrete choice experiment approach used in this study) could be adopted in order to elicit preference information. These include patient satisfaction surveys (Cohen, Forbes et al. 1996; Castle, Brown et al. 2005). However a limitation of this approach is the somewhat nebulous concept of 'patient satisfaction' (Williams 1994). Citizens' juries have also been used to obtain information on preferences (Mossialos and King 1999; Mooney and Blackwell 2004). Protagonists of citizens' juries suggest they allow for debate and reflection such that the preferences are better informed (Dolan, Cookson et al. 1999). However, opponents of this approach argue that empirical evidence indicates that juries are "chiefly concerned with non-rational persuasion, and because of this they are morally and democratically irrelevant." (Price 2000). By implication, therefore, far from helping to form more rounded preferences citizens' juries could serve to pervert them. Focus groups have also been used to obtain information on preferences (Bowie, Richardson et al. 1995; Bradley, Sweeney et al. 1999; Shaffer, Yebei et al. 2006). However, there is uncertainty about the optimal size of groups, or indeed

whether there is an optimal size, and what factors should determine this (Tang and Davis 1995). Focus groups may also be subject to the sort of criticisms made against citizens' juries (Price 2000), especially if the 'framing' of questions posed can sway the parameters of the debate.

Patient preferences are clearly important when considering new service models (Ryan, Scott et al. 2001; Jones, Berney et al. 2004) Although there is some information on patient preferences (Rodger, Gagne-Rodger et al. 2003), there remains a paucity of information on preferences for different models of healthcare provision. This analysis helps to bridge this gap, using DCEs to place monetary valuations upon characteristics of provision. Our analysis illustrates that both factors affecting health outcomes (e.g. speed of imaging diagnosis) and also factors which mainly affect the nature of the 'process' of care (e.g. continuity of care; hours of available provision; and the number of times patients have to go to hospital) are of importance to DVT patients. Moreover, respondents have clearly demonstrated a willingness to trade between differences in these attributes.

One of the implications of these findings is that studies which currently seek to establish whether patients prefer hospital outpatient care or community home based care (by simply comparing the two models) may miss a crucial point. Our DCE results, interpreted using a variety of approaches including estimates of Willingness to Pay (WTP) using marginal willingness to pay figures (MWTP), uptake rates, and compensating variation (CV), suggest that patient preferences for one model or the other are intimately related to the extent to which either model performs in terms of attributes that are considered important by the patient. Therefore, applying a mainstream WTP analysis which does not value the different attributes of hospital or community based provision would miss this crucial point, which the present study helps to draw attention to. Thus, patients' valuations may be related more to the degree to which prompt diagnosis or continuity of care can be provided by a model, rather than to whether or not the DVT service is provided in a

hospital or the community. It should be noted that in Leicester it was not common practice to use delayed repeat ultrasound scanning, so it made sense to include speed of diagnosis. However, delayed repeat ultrasound is standard practice in many protocols (Goodacre, Sampson et al. 2006) and is particularly relevant for below the knee DVT. The findings therefore, in relation to respondent valuation of quicker diagnosis, cannot be generalized to patients who are, or ought to be given, delayed repeat ultrasound on clinical grounds.

6. Conclusions

The implications of MWTP figures are unclear, if accurate they would suggest that patients do value certain attributes highly. For example, they value each 24 hour reduction in diagnostic delay at £115.73. This could be taken to imply it would be worth the UK NHS investing up to £115.73 for each 24 hour reduction in diagnostic delay. However, respondents had been told that 66% of patients in actual fact do not have a DVT, and unnecessary anticoagulation could be avoided as a result of quicker diagnosis. Therefore, in contexts in which more patients do actually have a DVT, valuation of quicker diagnosis may be lower. Also, to the extent that the diagnosis may be less than definitive, estimated WTP for this attribute may also be overstated.

The NHS operates within budget constraints, and whilst speeding up diagnosis may be cost-beneficial, other interventions may be more cost-beneficial and more of a priority. However, because our WTP findings are derived from a patient sample of respondents (rather than a general public one) strictly speaking they should not be used to inform the benefits side of a Cost-Benefit Analysis (CBA). This is because, ideally, CBA requires that a societal (general public) perspective should be adopted. It should be noted that one problem with applying this DCE to a general public sample of respondents might have been that the general public lack experience of the issues surrounding DVT service

provision. Therefore, their valuation of attributes may be less well informed than valuations obtained from DVT patients.

However, patients may place a higher valuation upon improvements in attributes of DVT service provision, than a member of the general public would, because they would benefit from improvements. So estimates of MWTP and WTP generated using this DCE data are likely to be higher than they would have been had they been obtained from a general public sample.

Moreover, even if speeding up diagnosis was cost-beneficial some radiology departments may have problems recruiting radiologists, so fast-tracking DVT patients may not be possible.

However, if there is scope to provide 'fast-track' diagnosis for suspected DVT at low cost, and staff are available, this 'fast-tracking' diagnosis for DVT may represent a highly cost-beneficial improvement in service provision. It is clear that in relative terms, the provision of a quicker diagnosis is valued far more (£115.73) by patients than reducing the number of required visits to hospital (valued at £17.12 per visit averted). This indicates that 'fast-tracking' diagnosis may be more important to patients than the context in which they receive a service for suspected DVT.

Our findings (from WTP analysis utilising MWTP figures, uptake rates, and the CV welfare measure) highlight the fact that the extent to which a service model performs in terms of meeting patients' preferences for defined characteristics of provision may be more important than the actual model of care adopted. Therefore I would argue that there is a need to move away from a sterile debate which compares one model with another, and instead to ask how well different models perform in routine service settings in terms of defined attributes.

Also because of the high valuation that patients appear to place upon reducing waiting time for diagnostic results (£115.73 per day) it may well be that the issue requiring most consideration is whether or not fast-tracked diagnosis can be made available for patients with suspected DVT. Moreover, if service quality in a routine setting is significantly worse than that observed in trials conducted in centres of excellence, there will be a divergence between reported efficacy and actual effectiveness. If this divergence affects attributes such as speed of diagnosis (valued at £115.73 per day delay avoided) then our analysis demonstrates that such differences may impose a large economic burden upon patients.

A former UK Health secretary Alan Milburn once argued that we needed to move away from a 'one size suits all' health service by facilitating patient choice. The analysis contained within this chapter illustrates the usefulness of DCE results in assisting health policy makers better facilitate patient choice. Firstly, the MWTP figures can help to give policy makers a feel for how much patients might value defined improvements in provision (e.g. speeding up diagnosis, improving the continuity of provision, reducing visits to hospital by providing services in the community rather than hospital, etc). MWTP figures also help policy makers see which attributes of service provision patients value most. They can then concentrate their efforts upon making improvements to those characteristics for which the ratio of benefits to costs is highest.

Secondly, if policy makers wish to offer a variety of service models to patients, then surveying patient preferences using DCEs can help predict likely uptake rates for the alternative models. Policy makers could thereby ensure that service providers are appropriately resourced to meet the demand for different models of care they might wish to provide. Thirdly, welfare measures including the CV measure can be used to assess whether expanding choice (by offering different models of provision) is associated with sufficiently large improvements in patient welfare (as measured by CV) to justify the expenditure associated with improving DVT provision.

Finally, the pilot analysis for this DCE was conducted over ten years ago, and DCE methods have moved on considerably since I undertook this work. If I were to undertake this study again I would undertake a thorough mainstream WTP analysis during piloting to establish an appropriate range for the monetary attribute. I would also aim to have an equal number of levels for all attributes if possible (in the interests of achieving level balance), and would therefore not resort to using some profiles twice from the design template in an attempt to help restore level balance. During piloting, I would fully explore (using qualitative interviews and 'think aloud' exercises) whether attributes might interact with each other in order to generate a design which catered for any interactions that might exist. I would also ensure that the DCE took on board all the developments in diagnosis of DVT which might have arisen in the last decade. In addition, I would also use 'cheap talk' in order to encourage respondents to factor in differences in the monetary attribute into their decision making, as this might considerably improve the reliability of estimated WTP. Furthermore, I would also take great care to ensure that the descriptor for the monetary attribute was as incentive compatible as possible in the interests of minimizing hypothetical bias. Finally, I would try to include a general public sample as well as a patient sample so that the WTP results could be used for the purposes of conducting a Societal Cost-Benefit Analysis.

In conclusion, I recognize that the reliability of the findings of the MWTP analysis, uptake analysis, and analysis of Compensating Variation (CV) contained in this chapter is compromised a little because of the aforementioned concerns. Nonetheless, the analysis in the chapter still helps to give a flavour of the potential that findings from MWTP analysis, uptake analysis, and welfare analysis (involving compensating variation) might have to inform policy if well conducted and not subject to bias.

7. **References**

Ageno, W. and A. G. Turpie (2005). "Clinical trials of deep vein thrombosis prophylaxis in medical patients." Clin Cornerstone **7**(4): 16-22.

Baltussen, R., E. Stolk, et al. (2006). "Towards a multi-criteria approach for priority setting: a application to Ghana." Health Economics **15 (July (7))**: 689-696.

Bowie, C., A. Richardson, et al. (1995). "Consulting the Public About Health-Service Priorities." British Medical Journal **311**(7013): 1155-1158.

Bradley, N., K. Sweeney, et al. (1999). "The health of their nation: how would citizens develop England's health strategy?" British Journal of General Practice **49**(447): 801-805.

Castle, N. G., J. Brown, et al. (2005). "Review of the literature on survey instruments used to collect data on hospital patients' perceptions of care." Health Services Research **40**(6): 1996-2017.

Clark, M. D., A. K. Gumber, et al. (2009). "Prioritising patients for renal transplantation? Analysis of patient preferences for kidney allocation according to ethnicity and gender." Diversity in Health and Care, 2009 **6**: 181-191.

Cohen, A. T., C. Hirst, et al. (2005). "Meta-analysis of trials comparing ximelagatran with low molecular weight heparin for prevention of venous thromboembolism after major orthopaedic surgery." Br J Surg **92**(11): 1335-1344.

Cohen, G., J. Forbes, et al. (1996). "Can different patient satisfaction survey methods yield consistent results? Comparison of three surveys." British Medical Journal **313**(7061): 841-844.

Dolan, P., R. Cookson, et al. (1999). "Effect of discussion and deliberation on the public's views of priority setting in health care: focus group study." BMJ **318**(7188): 916-919.

Douketis, J. D. (2005). "Treatment of deep vein thrombosis: what factors determine appropriate treatment?" Can Fam Physician **51**: 217-223.

Farrar, S., M. Ryan, et al. (2000). "Using discrete choice modelling in priority setting: an application to clinical service developments." Soc Sci Med **50**(1): 63-75.

Goodacre, S., F. Sampson, et al. (2006). "Measurement of the clinical and cost-effectiveness of non-invasive diagnostic testing strategies for deep vein thrombosis." Health Technol Assess **10**(15): 1-168, iii-iv.

Gutt, C. N., T. Oniu, et al. (2005). "Prophylaxis and treatment of deep vein thrombosis in general surgery." Am J Surg **189**(1): 14-22.

Huber, J. and K. Zwerina (1996). "The importance of utility balance in efficient choice designs." Journal of Marketing Research **33**(3): 307-317.

Jan, S., G. Mooney, et al. (2000). "The use of conjoint analysis to elicit community preferences in public health research: a case study of hospital services in South Australia." Aust N Z J Public Health **24**(1): 64-70.

Jones, I. R., L. Berney, et al. (2004). "Is patient involvement possible when decisions involve scarce resources? A qualitative study of decision-making in primary care." Soc Sci Med **59**(1): 93-102.

Lancaster, K. (1971). "Consumer demand: a new approach." (New York and London: Columbia University Press).

McIntosh, E. (2006). "Using discrete choice experiments within a cost-benefit analysis framework: some considerations." Pharmacoeconomics **24**(9): 855-868.

McIntosh, E., C. Donaldson, et al. (1999). "Recent advances in the methods of cost-benefit analysis in healthcare - Matching the art to the science." Pharmacoeconomics **15**(4): 357-367.

Michota, F. (2005). "Venous thromboembolism: epidemiology, characteristics, and consequences." Clin Cornerstone **7**(4): 8-15.

Montes, J., L. Gonzalez, et al. (2005). "[Home versus inpatient therapy for deep venous thrombosis. A cost-comparative analysis]." An Med Interna **22**(8): 369-372.

Mooney, G. H. and S. H. Blackwell (2004). "Whose health service is it anyway? Community values in healthcare." Medical Journal of Australia **180**(2): 76-78.

Morgan, A., P. Shackley, et al. (2000). "Quantifying patient preferences for out-of-hours primary care." J Health Serv Res Policy **5**(4): 214-218.

Mossialos, E. and D. King (1999). "Citizens and rationing: analysis of a European survey." Health Policy **49**(1-2): 75-135.

Nijkeuter, M., J. S. Ginsberg, et al. (2006). "Diagnosis of deep vein thrombosis and pulmonary embolism in pregnancy: a systematic review." J Thromb Haemost **4**(3): 496-500.

O'Brien, B., M. Levine, et al. (1999). "Economic evaluation of outpatient treatment with low-molecular-weight heparin for proximal vein thrombosis." Archives of Internal Medicine **159**(19): 2298-2304.

Petersen, P. (2005). "Ximelagatran--a promising new drug in thromboembolic disorders." Curr Pharm Des **11**(4): 527-538.

Pout, G., J. Wimperis, et al. (1999). "Nurse-led outpatient treatment of deep vein thrombosis." Nurs Stand **13**(19): 39-41.

Price, D. (2000). "Choices without reasons: citizens' juries and policy evaluation." J Med Ethics **26**(4): 272-276.

Rodger, M. A., C. Gagne-Rodger, et al. (2003). "The outpatient treatment of deep vein thrombosis delivers cost savings to patients and their families, compared to inpatient therapy." Thromb Res **112**(1-2): 13-18.

Ryan, M. (2004). "Deriving welfare measures in discrete choice experiments: a comment to Lancsar and Savage (1)." Health Econ **13**(9): 909-912; discussion 919-924.

Ryan, M. (2004). "Discrete choice experiments in health care." BMJ **328**(7436): 360-361.

Ryan, M. and K. Gerard (2003). "Using discrete choice experiments to value health care programmes: current practice and future research reflections." Appl Health Econ Health Policy **2**(1): 55-64.

Ryan, M. and J. Hughes (1997). "Using conjoint analysis to assess women' preferences for miscarriage management." Health Economics **6**: 261-273.

Ryan, M., D. A. Scott, et al. (2001). "Eliciting public preferences for healthcare: a systematic review of techniques." Health Technol Assess **5**(5): 1-186.

Ryan, M., V. Watson, et al. (2008). "Practical issues in conducting a Discrete Choice Experiment." In M. Ryan, K. Gerard and M. Amaya-Amaya (eds), Using Discrete Choice Experiments to Value Health and Health Care: 73 – 97.

Sassi, F., D. McDaid, et al. (2005). "Conjoint analysis of preferences for cardiac risk assessment in primary care." Int J Technol Assess Health Care **21**(2): 211-218.

Schraibman, I. G., A. A. Milne, et al. (2007). "Home verses in-patient treatment for deep vein thrombosis." Cochrane Database Systematic Review **3**: CD003076.

Schwappach, D. L. B. (2003). "Does it matter who you are or what you gain? An experimental study of preferences for resource allocation." Health Economics **12**: 255-267.

Shaffer, D. N., V. N. Yebei, et al. (2006). "Equitable treatment for HIV/AIDS clinical trial participants: a focus group study of patients, clinician researchers, and administrators in western Kenya." Journal of Medical Ethics **32**(1): 55-60.

Stannard, J. P., R. R. Lopez-Ben, et al. (2006). "Prophylaxis against deep-vein thrombosis following trauma: A prospective, randomized comparison of mechanical and pharmacologic prophylaxis." Journal of Bone and Joint Surgery-American Volume **88A**(2): 261-266.

Tang, K. C. and A. Davis (1995). "Critical factors in the determination of focus group size." Fam Pract **12**(4): 474-475.

Wichers, I. M., M. Di Nisio, et al. (2005). "Treatment of superficial vein thrombosis to prevent deep vein thrombosis and pulmonary embolism: a systematic review." Haematologica **90**(5): 672-677.

Williams, B. (1994). "Patient satisfaction: a valid concept?" Soc Sci Med **38**(4): 509-516.

Wooldridge, J. M. (2002). "Economic analysis of cross section and panel data." MIT press.

Chapter 3: Discrete choice experiments and willingness to pay analysis (WTP). An approach to assess the possible impact of hypothetical bias upon estimated WTP.

1. Introduction.

In this Chapter, the focus is upon exploring the impact of hypothetical bias on estimates of WTP obtained using DCEs when healthcare is free at the point of use. The DCE was one used to compare a 'One-Stop' outpatient model for women with menstrual disorders or other gynaecological problems with a new integrated care pathway service model.

In Chapter 1 (section E1.1) it was pointed out that one key concern about eliciting estimates of WTP has always been that estimates obtained using contingent valuation methods might be subject to hypothetical bias. It was pointed out that hypothetical bias refers to the difference between actual behaviour and hypothetical statements of value (Little and Berrens 2004; Mozumder and Berrens 2007). This form of bias therefore implies that the hypothetical nature of the preference elicitation has biased responses (Guzman and Kolstad 2007). However, it was further pointed out that there is reason to suppose that using DCEs to elicit WTP might make estimates less susceptible to some sources of hypothetical bias than using other forms of WTP analysis, including bias arising because of 'Yea saying', 'warm glow', embedding, and scope insensitivity (for further details see section E1.1).

In Chapter 2, I demonstrated how estimates of Marginal Willingness to Pay (MWTP), and estimates of Compensating Variation might be useful in informing health policy. Moreover, it can further be pointed out that if DCEs can be accurately applied to generate estimates of the general public's Willingness to Pay (WTP) for health interventions, then these WTP estimates

could be used to inform the benefits side of a Cost-Benefit Analysis (CBA) (McIntosh, Donaldson et al. 1999).

CBA might be preferred to Cost-Utility Analysis (CUA) to evaluate healthcare interventions if the interventions generate both process and health outcomes. This is because conventional CUA is ill equipped to value benefits which do not impact upon health outcomes. In contrast DCE studies can be designed to establish how much respondents value both health and process outcomes. If the information yielded about WTP obtained through DCEs is accurate, then this data could be used to derive welfare estimates (McIntosh and Ryan 2002). However, a number of theoretical and practical implementation issues remain surrounding the application of CBA (McIntosh and Ryan 2002; McIntosh 2006; Borghi 2007). Crucially if DCEs are to be used to calculate WTP and welfare estimates for CBA, then estimates of WTP must be accurate.

2. Background.

It was pointed out in Chapter 1 (section E1.1) that concerns that WTP estimates obtained using DCEs might be subject to hypothetical bias remain. DCEs have been equally applied when healthcare is free at the point of use (Propper 1994; Ryan and Hughes 1997; Bryan, Buxton et al. 1998; van der Pol and Cairns 1998; San Miguel, Ryan et al. 2000; Hall, Kenny et al. 2002; Taylor and Armour 2002; Weston and Fitzgerald 2004; Lloyd, McIntosh et al. 2005; Ryan, Diack et al. 2005) and when payment enters into decision making (Szeinbach, Mason et al. 1990; Stensrud, Sylvestre et al. 1997; Jan, Mooney et al. 2000; Telser and Zweifel 2002; Roux, Ubach et al. 2004; Hanson, McPake et al. 2005; van der Berg, Al et al. 2005). Therefore, not only might there be general problems with the robustness of using estimates of WTP using DCEs (Ratcliffe 2000; Ryan and Gerard 2003), but the WTP

estimates might also be more subject to hypothetical bias if DCEs are applied in a context in which healthcare provision is currently free at the point of use. For example, a respondent who does place a value on defined improvements may act strategically to conceal this (strategic bias). This is likely to be a particular problem with mainstream contingent valuation (CV) studies, although the potential for respondents to act strategically when answering DCE questions is not as great as in CV. Furthermore, respondents may have difficulty valuing health in monetary terms, and consider that some things are "not commensurable with monetary valuation." (Clark, Burgess et al. 2000). If respondents are faced with scenarios which appear unrealistic, lack of realism might also lead them to disregard dimensions of the choice set which they consider implausible. Respondents may also believe, for ideological or other reasons, that healthcare should be 'free at the point of use.' Finally, some respondents may interpret cost as something that can be ignored because in reality they don't face a real cost (Ratcliffe 2000).

If a proportion of respondents fail to factor in the monetary attribute into their decision making, then this would have the effect of making the estimated value of the coefficient upon the monetary attribute lower, which in its turn would have the effect of inflating estimated values of Marginal Willingness of Pay (MWTP). This is why in this chapter I present the methodology I have used to try to identify what proportion of respondents claim not to factor changes in the levels of the monetary attribute into their decision making, and then to evaluate what the impact of such behaviour might be upon estimated WTP.

As pointed out in Chapter 2, I began the pilot DCE analysis for the analysis contained in this chapter about 9 - 10 years ago. Consequently, I had not had the benefit of seeing the findings of some other pertinent analyses (Skjoldborg and Gyrd-Hansen 2003; Gyrd-Hansen and Skjoldborg 2008). The earlier analysis (Skjoldborg and Gyrd-Hansen 2003) suggested that

estimates of WTP may be sensitive both to the range specified for the monetary attribute and to the presence or absence of payment *per se*. In contrast, the later analysis (Gyrd-Hansen and Skjoldborg 2008) reached the conclusion that respondents might be more influenced by the presence or absence of a non-zero cost than by the level of cost indicated by the monetary attribute. Furthermore, the suggestion that respondents might have a lexicographic aversion to the "concept of paying" (Skjoldborg and Gyrd-Hansen 2003) may have particular salience in contexts in which healthcare is provided 'free at the point of use.' However, this is not something that the analysis presented in this chapter was designed to establish. In addition, the insight (Johnson, Mohamed et al. 2011) that respondents may deploy heuristics involving recoding costs into categories such as low, medium, or high, is again relatively new, and consequently did not feature as a consideration for the purposes of this analysis of a 'One-Stop' outpatient service model for women with menstrual disorders versus a new integrated care pathway service model.

Another potential source of hypothetical bias which the previous DCE analysis did not explore is 'framing effects' (Howard and Salkeld 2009). However, because the DCE analysis reported in this paper relates to screening for colorectal cancer, and in the present Chapter it relates to the management of menstrual disorders (or other miscellaneous gynaecological problems), the issue of 'framing effects' of this kind is likely to be less pertinent.

The focus of this chapter is upon the impact of hypothetical bias on estimates of WTP obtained using DCEs when healthcare is free at the point of use. The approach, as outlined, can be adopted even when it is not possible to compare stated preferences with actual or simulated market behaviour. Given that the potential for hypothetical bias is greatest with respect to goods and services which are free at the point of use (Ratcliffe 2000), an approach

which does not require external validation against actual or simulated market behaviour, is particularly useful.

This analysis suggests that if respondents behave in line with economic theory, they should value money. Money provides a medium of exchange which can be used to purchase goods and services conferring utility. Therefore, if a substantial proportion of respondents claim to disregard the monetary attribute ('cost to you') when making choices, this may be evidence of hypothetical bias. Moreover, if it can be demonstrated that this affects estimated WTP, it may be possible to make some assessment of whether hypothetical bias seriously influences estimated WTP (an internal test of validity). Note, we realize that differences in WTP could be partly be driven by the use of an inappropriate price vector for the monetary attribute, or it could be partly driven by the use of a monetary attribute which is not fully incentive compatible. However, we would argue that when you use hypothetical questions to value healthcare provision in a context in which healthcare is usually provided free at the point of use, it may not be possible to define a "correct" price vector for the monetary vector, or to always have an incentive compatible WTP question. Therefore the price vector for the monetary attribute, and questions used to elicit WTP might inevitably themselves engender some hypothetical bias.

Therefore when conducting the DCE discussed in this Chapter, which included a monetary attribute, I posed an additional question to establish whether respondents claimed to have factored in the monetary attribute into their decision-making. Subsequently, I went on to examine econometrically whether estimates of WTP varied according to how individuals had responded to this question. Any evidence of a statistically significant difference in WTP between those responding to this question in various ways would suggest that estimates of WTP / benefits may be biased.

3. Methods.

3.1. The policy issue to be assessed.

The DCE was applied to compare a One-Stop model for women with menstrual disorders or other Gynaecological problems, with a conventional approach. The One-Stop model involved women being referred by GPs to secondary care for outpatient consultation with hospital Consultants. This was compared to an integrated care pathway providing a bridge between primary and secondary care, allowing direct booking for investigations or treatments for patients who remained under the care of the GP (known as the Bridges Project). The model removed the traditional 'referral' between the healthcare sectors, allowing more involvement of GPs in the provision of specialized services. The DCE incorporated cost in the design; however, as the study population received NHS care 'free at the point of use', we assessed econometrically whether estimated WTP might be biased.

3.2. Piloting and designing the DCE questionnaire.

The DCE questionnaire was developed after extensive piloting. We established candidates for attributes based on insights from the medical literature, clinicians, and discussions with patients about whether they could relate to the range of attributes and levels we specified in the pilot questionnaire (which had been informed by insights from the medical literature and discussions with a small number of clinicians). We then conducted a ranking exercise using a sample of 60 patients to identify key attributes for inclusion in the pilot DCE questionnaire. We invited the patients to suggest other attributes and levels in one to one interviews. Any plausible suggested attributes would then be added to a list of candidate attributes which

could be included in the final questionnaire. This list would then be updated after every interview, and shown to other patients in later interviews so that they could rank attributes on the list (with associated levels listed next to them), in order of priority to them. We also conducted a ranking exercise in relation to the choice of descriptor for the monetary attribute. This involved presenting respondents with a range of possible ways of describing the monetary attribute, in order to establish which descriptor respondents felt most comfortable with (in order to inform the choice of monetary descriptor for the final questionnaire).

A main effects pilot questionnaire was designed using SPEED version 2.1 (Bradley 1991) using a fixed comparator with levels set to correspond to our perception of actual levels in Leicestershire for a 'typical' patient. The pilot questionnaire was completed by 60 respondents (30 patients had a questionnaire with a WTP ceiling of £125, and 30 with a ceiling of £250). We used two different potential ranges for the monetary attribute because we wanted to establish what range for the monetary attribute might be most appropriate for the final questionnaire. A researcher observed respondents as they completed the questionnaire. Regression results from the pilot DCE, together with researchers' observations, informed the final selection of attributes and levels.

3.3. The final selection of attributes and levels, and the design for final questionnaire

The final DCE questionnaire contained the attributes and levels described in table 1. The monetary descriptor 'Cost to you' was the one that most respondents felt comfortable with. Table 1 also indicates the variable names for attributes used for regression analyses. Having determined attributes and levels, we used SPEED again (Bradley 1991) to generate choice scenarios for a final main effects design. For the purposes of this DCE we used a fixed comparator and the 16 choices generated by SPEED were just paired off alongside against

this same fixed comparator for every choice specified in the design (see questionnaires in Appendix C). The 16 choices generated by SPEED were compared with a fixed comparator corresponding to the status quo in Leicestershire. The levels for the fixed comparator related to what we perceived the levels of the attributes would be for the average respondent under the traditional model of Consultant-led care in hospitals in Leicester (where this research was conducted). The fixed comparator assumed a 1 day wait for test results; no cost to you; care via a male consultant; a 12 week wait for an appointment; and getting to see the same doctor about half the time. All the Gynaecology hospital Consultants in Leicester at the time were men (so the attribute level relating to gender of the doctor was set to 'Male'). With Consultant-led care the average waiting time for test results was 1 day, and there was no charge for this NHS provision (so 'Cost to you' would = £0). The waiting time to see a consultant at the time was 12 weeks (in contrast if the provision was provided by GPs trained to provide this provision, it might be quicker). Moreover, under 'Consultant-led' care, patients could expect to see the same doctor about half of the time.

We adopted a 'fixed comparator' because some of the respondents had not got particularly high cognitive ability according to the medical researcher undertaking the pilot analysis. We therefore considered that the respondents would find it easier to answer the questions, if they had a consistent benchmark (the status quo option, option A) against which to evaluate alternative models of service provision. After generating the choice options using SPEED, I used SPSS to ensure that there was no collinearity between attribute levels, as a final design check.

The 16 choices which we used for the purposes of data analysis were divided, so any one respondent would only face half of them. Also, because of concerns that the ordering of questionnaire attributes (in the list of attributes faced by respondents), might affect

respondents' valuation of them we rotated the order of attributes. This meant that we had 12 final DCE questionnaires for women with 'Period Problems' and the 'cost to you' attribute and attributes were rotated so that the 'cost to you' attribute was 1st in the list for 2 versions, 2nd in the list for 2 versions, 3rd in the list for 2 versions, 4th in the list for 2 versions, 5th in the list for 2 versions, and 6th for 2 versions. We did not have the intention of testing to see whether estimated WTP varied according to the positioning of the monetary attribute. However, we reasoned that if the positioning of the monetary attribute was varied in this consistent manner, then estimated WTP would not be biased due to attribute ordering effects. We also conducted a later DCE analysis upon women with a range of other Gynaecological problems (see chapter 4) in which about 1 / 3rd of the respondents faced the same DCE scenarios with a 'cost to you' monetary descriptor. For the purposes of this later DCE analysis about 2 / 3rd of questionnaire respondents answered DCE questions which were otherwise the same except for the fact that the descriptor for the monetary variable was instead either 'Amount lost' or 'Willingness to Pay.' (these respondents were not included in the analysis in this chapter). However, we included within the data analysis detailed in this chapter (chapter 3) those respondents answering a 'Gynaecology healthcare' survey' who answered a questionnaire with a 'Cost to you' monetary descriptor. Once again there were 12 variants of this 'cost to you' 'Gynaecology healthcare' survey' questionnaire. This was because we rotated the order of attributes (to avoid ordering effects), and each respondent only faced half of the scenarios emanating from the SPEED design template.

Table 1: Details of attributes and levels.

Variable	Variable / Attributes	Levels
typd	The type of doctor you see	GP or Consultant
gend	The sex of the doctor you see	Male or Female
wait	Time spent waiting for an appointment to see the doctor (either the GP or the consultant)	1 day, 4 days, 6 weeks, 12 weeks
cont1 cont2	How often you get to see the same doctor	All of the time, half of the time. Half of the time, none of the time.
res	How long you have to wait for test results	1 day, 2 days, 2 weeks, 4 weeks
cty	Cost to you (i.e. perhaps because of absence from work or travel costs – Respondents were told to: Please assume you would lose this amount of money even if in reality you would not).	£0, £25, £75, £125

Some attributes and levels require further explanation. During piloting, and as a consequence of interviews with patients, we found that over 40% of women had a preference for a female rather than a male doctor (no one expressed a preference for a male rather than female doctor). We therefore included an attribute to enable us to value any such preferences for gender of the doctor. This was important, because at the study sites, all Gynaecology hospital consultants were male. Therefore, women could not select a female hospital doctor. In contrast, within primary care there was more scope for patients to select a female doctor.

The range for "time waiting for an appointment to see the doctor (either GP or Consultant)" was selected because a ceiling of 12 weeks corresponded to the maximum waiting time experienced by patients in the locality for a referral to a Consultant Gynaecologist. GP appointments can generally be arranged within a matter of days. However, GPs who provide Gynaecological support under the 'Bridges' model take on extra responsibility. Therefore patients who are non-emergency cases might be offered a pre-arranged appointment in several weeks' time. Some GPs might offer more prompt provision, but that would be determined by local factors. Certain GPs could offer patients either a referral to a hospital Consultant, or a practice based attendance, perhaps in several weeks. So the choice of levels for that attribute is not unrealistic. The levels for the "waiting for test results attribute", was set with reference to the minimum and maximum anticipated turn-around time for a variety of tests that might be required in this patient group. Finally, we decided to use the lower ceiling of £125 for the WTP attribute, as trading between the monetary and other attributes happened more often with this ceiling on WTP in the pilot.

3.4. Other questions included in the questionnaire.

We requested information on the socioeconomic status of respondents (see questionnaires in Appendix C). In addition after the pairwise choices, we also asked whether respondents factored in the 'cost to you' attribute when making choices:

Did differences in the amount of 'cost to you' for options A and B influence your choices?

❑ Yes ❑ No ❑ Sometimes

(Question 1)

One limitation of posing a question about the influence of the monetary attribute upon respondent choice is of course that respondents might take 'cost to you' into account when answering some questions but not others. This was why it was essential that respondents had the option to indicate that they 'Sometimes' took 'cost to you' into account.

Respondents had also been told in the questionnaire that average gross household income in Leicestershire was approximately £25,000. Respondents could then indicate whether they had above average, average, or below average income.

3.5. Data analysis and hypotheses tested.

3.6. Hypothesis that a basic model is adequate:

If DCE responses are not affected by patients' responses to question 1, then a random effects Probit model (model 1) using the variables defined in table 1 for all respondents could be used, and results would not be biased because of the failure of some respondents to

consider the ‘cost to you’ attribute.

$$Y_{ij} = \alpha_0 con + \alpha_1 typd_{ij} + \alpha_2 gend_{ij} + \alpha_3 wait_{ij} + \alpha_4 cont1_{ij} + \alpha_5 cont2_{ij} + \alpha_6 res_{ij} + \alpha_7 cty_{ij} + \mu_i + \xi_{ij}$$

(Model 1)

Here Y_{ij} is the binary dependent variable, from individuals i = 1…m, for observations j = 1…n_i. The number of observations n_i varies because the i individuals do not all complete every pairwise choice (a minority of respondents don’t answer all choices). The term μ_i is the random effects error term (which allows for multiple responses from i respondents), and ξ_{ij} is the standard Probit error term for individuals i, for j observations. Given that the DCE involved the use of a fixed comparator, and we set the levels for the fixed comparator equivalent to what we thought the average respondent would experience, under the traditional Consultant-led’ care model, then the constant if it is significant, might pick up whether or not respondents favoured the ‘status quo’ package bundle of attributes (option A) or not (so it could pick up an ‘experience effect.’)

3.7. Hypothesis that a segmented model is required:

If results are sensitive to how respondents respond to question 1, then model 2 may provide more accurate estimates. Variables for all the models are as defined in table 2.

Table 2: Variables defined.

Attributes	**Variables for the base group who claim they consider CTY when choosing or they 'sometimes' consider CTY when choosing**	**Dummy variables for those who say 'no' they don't consider CTY when choosing.**
The type of doctor you see	typd	Dntypd
The sex of the doctor you see	gend	Dngend
Time spent waiting for an appointment to see the doctor (either the GP or the consultant)	wait	Dnwait
How often you get to see the same doctor All of the time rather than half of the time Half of the time rather than none of the time	 cont1 cont2	 Dncont1 Dncont2
How long you have to wait for test results	res	Dnres
Cost to you (i.e. perhaps because of absence from work or travel costs – Respondents were told to: <u>Please assume you would lose this amount of money even if in reality you would not).</u>	cty	Dncty

Model 2 runs a model equivalent to model 1, but uses an additive dummy plus interaction dummy variables (variables with a 'Dn' prefix) to value differences in attributes differently for the sample (66 / 180 [36.7%]) of respondents who claimed they do not consider 'Cost to you' when choosing.

$$Y_{ij} = \alpha_0+\alpha_1 typd_{ij}+\alpha_2 gend_{ij}+\alpha_3 wait_{ij}+\alpha_4 cont1_{ij}+\alpha_5 cont2_{ij}+\alpha_6 res_{ij}+\alpha_7 cty_{ij} + \alpha_8 Dn_{ij}+\alpha_9 Dn_{ij} typd_{ij} +\alpha_{10} Dn_{ij} gend_{ij}+\alpha_{11} Dn_{ij} wait_{ij}+\alpha_{12} Dn_{ij} cont1_{ij} +\alpha_{13} Dn_{ij} cont2_{ij}+\alpha_{14} Dn_{ij} res_{ij}+\alpha_{15} Dn_{ij} cty_{ij}+\mu_i+\xi_{ij}$$

(Model 2)

Respondents who truthfully respond that either 'Yes' they did take differences in the monetary attribute into account, or they 'Sometimes' did, may be behaving in an unbiased manner. Indeed since some respondent choices involved there not being a 'Cost to you' for both option A and B respondents, and so rationally respondents might ignore differences in 'Cost to you' for such options. Whilst of course if respondents answered 'Yes' they did take differences in the monetary attribute into account, this is reassuring.

However, if respondents truthfully answer that 'No' they did not take differences in the value of the monetary attribute into account then this may indicate that they are disregarding differences in 'cost to you' as a result of hypothetical bias. Therefore establishing separate estimates of coefficients for the group of respondents who answer 'No' may allow us to establish how this might impact upon estimated WTP.

3.8. Likelihood ratio test.

A likelihood ratio test was used to compare the unrestricted model, model 2 (using the log-likelihood from model 2), with the restricted model (using the log likelihood from model 1).

4. Results.

4.1. Patient sample profile:

Most of the patients (n = 121) presented with menstrual disorder (and answered a questionnaire with a heading relating to ‘Period Problems’, but others (n=59) presented with other Gynaecological problems, and responded to a questionnaire with the same DCE choices, but with a different front cover on it. Details of some key patient sample characteristics are outlined in table 3.

Table 3: Respondents self reported age, household income, and response to the question about whether they consider 'cost to you'.

Age range of patients	**Patients with Menstrual disorder (n=121)**	**Patients with other gynaecological symptoms (n=59)**	**Combined group (n=180)**
20-30	2 / 121 (1.7%)	10 / 59 (16.9%)	12 / 180 (6.7%)
31-40	14 / 121 (11.6%)	21 / 59 (35.6%)	35 / 180 (19.4%)
41-50	72 / 121 (59.5%)	13 / 59 (22.0%)	85 / 180 (47.2%)
51-60	30 / 121 (24.8%)	8 / 59 (13.6%)	38 / 180 (21.1%)
60+	0 / 121	2 / 59 (3.4%)	2 / 180 (1.1%)
No response	3 / 121 (2.5%)	5 / 59 (8.5%)	8 / 180 (4.4%)
Average age of those responding	44.87 years	39.47 years	43.10 years
Household income of patients.			
Above average	53 / 121 (43.8%)	12 / 59 (20.3%)	65 / 180 (36.1%)
About average	26 / 121 (21.5%)	23 / 59 (39.0%)	49 / 180 (27.2%)
Below average	37 / 121 (30.6%)	17 / 59 (28.8%)	54 / 180 (30%)
No response	5 / 121 (4.1%)	7 / 59 (11.9%)	12 / 180 (6.7%)
Do respondents report they consider the 'the cost to you attributes' when making decisions			
Yes	22 / 121 (18.2%)	18 / 59 (30.5%)	40 / 180 (22.2%
No	48 / 121 (39.7%)	18 / 59 (30.5%)	66 / 180 (36.7%)
Sometimes	51 / 121 (42.1%)	23 / 59 (39.0%)	74 / 180 (41.1%)

4.2. Econometric analysis.

Random effects probit was conducted using STATA v. 9.2. After conducting the econometric analysis we calculated WTP for attributes, and associated asymptotic confidence intervals using the Delta method (Wooldridge 2002). The Delta method was implemented using the STATA non-linear confidence interval function, and Wald tests using the STATA non-linear test command.

4.3. Comparison of restricted vs. unrestricted models using a likelihood ratio (LR) test.

We obtained the likelihood ratios for model 1 (the restricted model) and model 2. The figures for model 1 and 2 respectively are 1097.5 and 1058.2 respectively. This suggests a likelihood ratio test value of 78.6. This compares with the critical value for 8 degrees of freedom of 15.5. Using an LR test, we therefore reject the null hypothesis that model 1 (the restricted model) provides an adequate characterization of the underlying data, compared to model 2. This suggests that WTP results are jointly significantly different amongst the group of respondents who indicated that 'No' they did not take differences in the monetary attribute into account.

4.4. Findings from the econometric models.

The variables are as defined in table 1, and econometric results for models 1 and 2 are in table 4.

Table 4: Econometric results for models 1, and 2.

Attribute	Model 1 (Pooled model).	Model 2 (Coefficients for the base group who claim 'yes' they do, or they 'sometimes' take differences in CTY into consideration).	Model 2 (Coefficients on dummy variables for the group who claim 'no' they do not take differences in CTY into consideration).
Intercept	-0.092 (p =0.058)	-0.048 (p=0.512)	-0.079 (p=0.516)
typd	0.351 (p = 0.000)	0.316 (p = 0.000)	0.143 (p=0.296)
gend	0.238 (p = 0.000)	0.149 (p=0.062)	0.251 (p=0.070)
wait	-0.012 (p = 0.000)	-0.010 (p=0.000)	-0.005 (p=0.003)
cont1	0.301 (p = 0.000)	0.223 (p=0.011)	0.207 (p=0.165)
cont2	0.290 (p = 0.001)	0.296 (p=0.005)	-0.004 (p=0.983)
res	-0.026 (p = 0.000)	-0.029 (p=0.000)	0.005 (p=0.362)
cty	-0.006 (p = 0.000)	-0.008 (p=0.000)	-0.004 (p=0.004)
Sample size / observations	180 patients in this sample	114 / 180 patients (63.3%) in this category.	66 / 180 patients (36.7%) in this category
Likelihood ratio test model 1 (restricted) vs. model 2.			78.6 is the actual value, the critical value for 8 degrees of freedom = 15.5: Jointly significant difference.
Mc Faddens R-squared	0.1873		0.2163
Percentage of actual values predicted.	74.1%		74.9%

4.5. Calculating WTP.

We assume a 'state of the world' model in which WTP for one alternative is to be considered (Ryan 2004). Table 5 details the formulas underpinning the functional form for the WTP point estimates.

Table 5: Coefficients used to derive estimates of WTP.

Variable	Model 1 – WTP estimates: Pooled model (using coefficients from model 1 results)	Model 2 – WTP estimates: Do or sometimes factor in 'Cost to you' (using coefficients from model 2 results)	Model 2 - WTP estimates: Don't factor in 'Cost to you' (using coefficients from model 2 results)
Type of doctor	$(\alpha_1) / -(\alpha_7)$	$(\alpha_1) / -(\alpha_7)$	$(\alpha_1+\alpha_9) / -(\alpha_7+\alpha_{15})$
Gender of doctor	$(\alpha_2) / -(\alpha_7)$	$(\alpha_2) / -(\alpha_7)$	$(\alpha_2+\alpha_{10}) / -(\alpha_7+\alpha_{15})$
Extra days waiting to see doctor	$(\alpha_3) / -(\alpha_7)$	$(\alpha_3) / -(\alpha_7)$	$(\alpha_3+\alpha_{11}) / -(\alpha_7+\alpha_{15})$
Continuity (Same doctor all the time not half of the time)	$(\alpha_4) / -(\alpha_7)$	$(\alpha_4) / -(\alpha_7)$	$(\alpha_4+\alpha_{12}) / -(\alpha_7+\alpha_{15})$
Continuity (Same doctor all the time not half of the time)	$(\alpha_5) / -(\alpha_7)$	$(\alpha_5) / -(\alpha_7)$	$(\alpha_5+\alpha_{13}) / -(\alpha_7+\alpha_{15})$
Extra days waiting for test results	$(\alpha_6) / -(\alpha_7)$	$(\alpha_6) / -(\alpha_7)$	$(\alpha_6+\alpha_{14}) / -(\alpha_7+\alpha_{15})$

4.6. WTP results.

For illustrative purposes we interpret the results from model 1 (table 6) assuming they are unbiased. The results in table 4 suggest that the constant term is not statistically significant at the 5% level. So at the 5% significance level there is no statistically significant evidence of an 'experience effect' i.e. respondents either favouring or disliking option A, which was equivalent to the bundle of characteristics the average respondent currently received (i.e. this questionnaire was used upon patients currently receiving Consultant-led care Gynaecological care before the new GP led provision was introduced). Given the coding we ascribed to the variables, the findings suggest that taking the sample overall respondents value having a Consultant rather than a GP at £58.13 (£32.87 : £83.39). A female rather than a male doctor is valued at £39.45 (£20.42 : £58.49). Avoiding an extra day spent waiting to see the doctor is valued at £1.92 (£1.46: £2.38). Getting to see the same doctor all of the time rather than half of the time is valued at £49.84 (£28.85 : £70.83); whilst seeing the same doctor half of the time rather than none of the time is valued at £48.02 (£17.07 : £78.97). Avoiding having to wait an extra day for test results is associated with a WTP of £4.23 (£2.78: £5.67). It follows that the interpretation of the results for models 2 is similar (because all the results are of the same sign). However, as table 6 shows WTP and corresponding confidence intervals does vary comparing the pooled model (model 1) with the unrestricted model (model 2).

Table 6: Estimated WTP for models 1 and 2.

Variable	Model 1 Pooled model (all responders)	Model 2 Do or sometimes factor in 'Cost to you' group	Model 2 Don't factor in 'Cost to you' group
Type of doctor	£58.13** (£32.87 : £83.39)	£40.67** (£17.90 : £63.43)	£131.89* (£20.89 : £242.88)
Gender of doctor	£39.45** (£20.42 : £58.49)	£19.15* (£0.21: £38.10)	£.114.92** (£37.32: £192.53)
Extra days waiting to see doctor	-£1.92** (-£1.46 : -£2.38)	-£1.34** (-£0.99 : -£1.67)	-£4.52** (-£1.56 : -£7.48)
Continuity (Same doctor all the time not half of the time)	£49.84** (£28.85 : £70.83)	£28.73** (£7.96 : £49.49)	£123.61** (£40.18 : £207.04)
Continuity (Same doctor all of the time not half of the time)	£48.02** (£17.07 : £78.97)	£38.13* (£8.60 : £67.66)	£84.05 (-£21.58 : £189.68)
Extra days awaiting test results	-£4.23** (-£2.78 : -£5.67)	-£3.67** (-£2.34 : -£5.00)	-£6.74* (-£1.05 : -£12.43)

* denotes significant at the 5% level; **denotes significant at the 1% level; figures in brackets are 95% confidence intervals.

It is quite clear just from looking at estimated of WTP (table 6) that estimated WTP seems to vary considerably between the pooled model (model 1), and a model distinguishing between those who do and who do not factor 'Cost to you' into their decision making (model 2).

We conducted Wald tests to assess whether estimates of WTP generated using models 1 and 2 varied. Details of Wald tests conducted, and results are tabulated in table 7. These Wald test results suggest that WTP differs significantly at the 5% level for 3 / 6 attributes in model 2 (comparing those who say they do or 'sometimes' factor 'cost to you' into their decision making and those who said they did not).

Table 7: Wald tests of restrictions on WTP.

Wald Test	**Do take monetary descriptor into account group vs. 'sometimes' take monetary descriptor into account group- Restriction tested** **(Model 2)**	**Results** **n = 180**
1	$\alpha_2 / -(\alpha_8) =$ $(\alpha_2 + \alpha_{10}) / -(\alpha_8 + \alpha_{16})$	p = 0. 1146
2	$\alpha_3 / -(\alpha_8) =$ $(\alpha_3 + \alpha_{11}) / -(\alpha_8 + \alpha_{16})$	p = 0.0188*
3	$\alpha_4 / -(\alpha_8) =$ $(\alpha_4 + \alpha_{12}) / -(\alpha_8 + \alpha_{16})$	p = 0.0361*
4	$\alpha_5 / -(\alpha_8) =$ $(\alpha_5 + \alpha_{13}) / -(\alpha_8 + \alpha_{16})$	p = 0.0305*
5	$\alpha_6 / -(\alpha_8) =$ $(\alpha_6 + \alpha_{14}) / -(\alpha_8 + \alpha_{16})$	p = 0.4117
6	$\alpha_7 / -(\alpha_8) =$ $(\alpha_7 + \alpha_{15}) / -(\alpha_8 + \alpha_{16})$	p = 0.3036

5. Discussion.

In this Chapter, I wanted to establish whether people are factoring in the actual level of 'cost to you' when making a decision about whether to choose option A or B. The fact that 66 / 180 (36.7%) of respondents (table 3) say 'No' is of concern. If respondents behaved in line with economic theory, we would expect them to value money. Therefore, a failure to take CTY into account completely may be indicative of hypothetical bias. We concede that this bias might be reduced / prevented if we could ensure that a more appropriate price vector was used for the monetary attribute, perhaps by deploying more advanced methods to set the price vector for the monetary attribute (i.e. use of a payment card during piloting), or if we could find a WTP question which was more incentive compatible. Moreover, if respondents do not factor CTY into their decision making, the coefficient is likely to be underestimated for the sample overall, resulting in inflated estimates of WTP in the baseline analysis (i.e. model 1).

One surprising result (table 4) is that the coefficient on 'cost to you' remains highly significant, even in the group of respondents who claimed they did not take this into account when making pairwise choices. If respondents had reported accurately whether or not they had taken differences in 'cost to you' into account, then we would have expected that this coefficient would prove insignificant for the group who claimed they did not take differences in 'cost to you' into account. Part of the reason for this may be due to some respondents answering 'No', because they <u>usually</u> do not take differences in the monetary attribute into account (but in reality, for a few cases, they did take such differences into account). We had hoped that such respondents would have ticked the 'Sometimes' option instead. If this response did differ depending on the case, it may be worth posing a question about whether differences in the monetary attribute affected their choice after <u>each</u> DCE question.

Given that there is some evidence (Ozdemir, Johnson et al. 2009) that 'cheap talk' may be required to get respondents to focus on the actual levels of the monetary attribute when making choices, it is possible that posing a question about how the monetary attribute influenced their decision after each DCE choice may simply have a similar effect to 'cheap talk' in terms of making respondents more inclined to factor-in differences in the monetary attribute into their decision making. This might be useful if our aim is simply to make WTP estimates less prone to bias. But it would not help with addressing the question we planned to address in this Chapter; whether DCE questions involving a monetary attribute have a general susceptibility to hypothetical bias because respondents fail to factor differences in the levels of the monetary attribute into their decision-making.

In the DCE presented here, the coefficient on the monetary attribute remains significant and negative, but is smaller amongst respondents who claimed they did not take differences in 'cost to you' into account. Tables 6 details these WTP results. The Wald tests (table 7) provide evidence that for model 2 there are statistically significant differences in WTP with respect to 3 / 6 (50%) of variables (gend, wait, and continuity of provision) between those who do and do not factor in CTY into their decision making. For model 2 this suggests that the inclusion of a sample of respondents who claim not to take differences in CTY into account is associated with a statistically significant difference in WTP for 3 / 6 (50%) of the non-monetary attributes.

It is interesting to note that, whilst the group that reports it does not consider the monetary attribute is associated with a lower weighting on the CTY coefficient than the base group, WTP remains significant in this group of respondents, in model 2. It is possible that a proportion of respondents may not like to admit to taking money into account when valuing

healthcare provision, resulting in an overestimate of those disregarding the CTY attribute. If this is the case it suggests our methodology might be open to question, because respondents cannot be relied upon to be honest (which is the main limitation of this approach). This inevitably means the policy implications of the DCE analysis in its current form would be open to question as well

6. Conclusions.

If DCEs are applied when healthcare is free at the point of use, there may be a lack of incentive to state preferences accurately, leading to the problem of hypothetical bias. In our study we found that 36.7% of respondents reported that differences in 'cost to you' did not influence their choices, and in econometric results we discovered that the inclusion of these respondents served to inflate estimated WTP. We consider that this is an extreme result which could have been avoided by firstly adopting more rigorous approaches to establishing an appropriate price vector for CTY, and then if we could establishing a more incentive compatible WTP question. That said, sometimes it will not be possible to establish a 'definitive' price vector, especially when healthcare provision is free at the point of use, no matter how rigorous our attempts are to identify what it should be. Nor are we ever likely to be able to generate WTP questions which are completely incentive compatible for healthcare if it is free at the point of use. However, it has also been suggested to me that, because we opted for the lower of the two ranges from the pilot analysis for the monetary attribute (with a ceiling on WTP of £125, not £250), some respondents may simply have felt that the levels of 'cost to you' were insufficiently high to be worthy of consideration.

For this reason, if I were to repeat a similar analysis I would both want to use mainstream WTP analysis to establish respondents' maximum WTP in a pilot sample (and use this

information to establish a more rigorously determined set of levels for the monetary attribute), and I would also allow for the use of 2 or more payment ceilings upon the levels for the monetary attribute (using different versions of the questionnaire with different levels for the monetary attribute). This would enable me to establish whether decisions to take differences in the monetary attribute into account might be related to the appropriateness / inappropriateness of the levels set for this attribute. Ideally, I would also want to use qualitative approaches such as 'think aloud' exercises (Ryan, Watson et al. 2009) to establish whether respondents who claimed not to take differences in the monetary attribute into account, actually appeared to be behaving in this way when they 'thought aloud' about the choices they made. It may also be worth using qualitative interviews to probe which factors influence respondents who claim not to take differences in the monetary attribute into account. For example, is this choice related to a general feeling amongst such respondents that healthcare should be available free at the point of use; or is it related to a general insensitivity to changes in the levels of the monetary attribute; or is it related to a feeling that in reality a monetary attribute would not apply so differences in it should be ignored?

The literature suggests that it is far from clear that responses to hypothetical questions in DCEs will necessarily elicit a carefully considered response (Braden, Kolstad et al. 1991). As a result, some respondents may fail to factor the monetary attribute into their decision making. This may mean that using 'cheap talk' (Ozdemir, Johnson et al. 2009) to reduce the number of respondents not taking a monetary attribute into account should be considered for piloting. Response patterns might be different in the context of 'thinking aloud' exercises than they would be if DCEs were conducted without a 'think aloud' exercise. Indeed, there is some evidence to suggest that unless 'cheap talk' is used to ensure respondents think about the different levels of a monetary attribute, their responses tend to be insensitive to changes in the levels of the monetary attribute (Ozdemir, Johnson et al. 2009). This suggests that the

use of ‘cheap-talk’ might be required to more accurately obtain estimates of WTP. Although ‘cheap talk’ can help to ensure that the cost function is linear (Ozdemir, Johnson et al. 2009), if it is the case that some respondents deploy heuristics to recode costs into categories such as low, medium, or high (Johnson, Mohamed et al. 2011), or even worse have a lexicographic aversion to the concept of paying (Skjoldborg and Gyrd-Hansen 2003), then estimates of WTP obtained using DCE might inevitably be subject to some degree of bias. Such bias is especially likely if DCEs are applied in contexts in which healthcare is free at the point of use.

Research also suggests that preferences may be subject to a ‘learning curve’, such that later responses in a DCE might be a better indicator of preferences than earlier responses (Johnson and Desvousges 1997; Carlsson and Martinsson 2001). This fits with evidence that estimates of WTP might be affected by the number of DCE choices that respondents face (Bech, Kjaer et al. 2011). This evidence is compatible with the view that respondents might use decision making heuristics (Lloyd 2003), and develop their decision making heuristics via ongoing learning. If respondents use simplifying heuristics (Lloyd 2003) they will ignore much of the evidence they are presented with, and adopt simplifying decision rules. It has been suggested that the more complex choices become in relation to options and variability within options, the less people are likely to engage in compensatory decision-making (Payne, Bettman et al. 1993). There is also evidence that more complex choice experiments are associated with more complexity-induced choice inconsistency (DeShazo and Fermo 2002); whilst other evidence suggests that ‘fatigue-effects’ may influence responses if more than 10 or so choice comparisons are presented, such that the survey instrument itself exerts influences upon choice, which would not be present in actual choice situations (Bradley and Daly 1994).

The use of simplifying heuristics by respondents might result in a failure to consider the monetary attribute. It may also explain some evidence which indicates (Kjaer, Bech et al. 2006) that the positioning of the price attribute within DCEs can affect price sensitivity. The current DCE analysis was designed to avoid this source of bias by rotating the positioning of the monetary attribute across different versions of the questionnaire, which were then distributed in equal numbers. Also, if the range set for the monetary attribute is inappropriate, there is evidence that this can affect WTP results (Ratcliffe 2000; Ryan and Wordsworth 2000; Skjoldborg and Gyrd-Hansen 2003). This might be reflected in respondents failing to factor in the monetary attribute into their decision-making.

This is why, if I were repeating a similar analysis again, I would use appropriate mainstream WTP analysis (not using DCEs but probably using payment cards) during the pilot exercise in order to more rigorously establish a sensible range for the monetary attribute. I would also avoid using a descriptor for the monetary attribute like 'cost to you.' I used this descriptor because, in the pilot exercise, respondents preferred it over and above every other option. However, this descriptor may encourage cost based responses (i.e. people valuing things in terms of what they consider to be a reasonable cost rather than its actual value). In fact the analysis presented in the next chapter (Chapter 4) suggests that the use of other descriptors as an alternative to 'cost to you' (including 'Amount lost' and 'Willingness to Pay') is not associated with a statistically significant difference in estimated WTP. Therefore use of the descriptor 'cost to you' may not have resulted in much bias due to cost based responses. That said, the fact that we told respondents when explaining 'Cost to you' in the preamble to "Please assume you would lose this amount even if you would not" meant that the way the monetary attribute was presented was not particularly incentive compatible. Therefore, if I were to undertake a similar analysis again I would do whatever I could to ensure a more incentive compatible monetary attribute was presented to respondents. I would also conduct

a final analysis which allowed the range of the monetary attribute to vary, so that I could identify any possible influence of the levels used for the monetary attribute upon whether or not respondents take monetary differences into account. Moreover, I would also use qualitative research, including ‘think-aloud’ exercises, to gain more insight into respondent motivations and behavior.

7. References.

Bech, M., T. Kjaer, et al. (2011). "Does the Number of Choice Sets Matter? Results from a Web Survey Applying a Discrete Choice Experiment." Health Econ **20**(3): 273-286.

Borghi, J. (2007). "Aggregation rules for cost-benefit analysis: A health economics perspective." Health Economics **Early View: Published on line**(DOI: 10.1002 / hec. 1304).

Braden, J. B., C. D. Kolstad, et al. (1991). ""Introduction." Measuring demand for enviromental quality." in Braden J. B. and Kolstad, C. D. (eds.) (Amsterdam: North-Holland).

Bradley, M. (1991). "Users manual for SPEED version 2.1 stated preference editor and designer." The Hague: Hague Consulting Group.

Bradley, M. and A. Daly (1994). "Use of the Logit Scaling Approach to Test for Rank-Order and Fatigue Effects in Stated Preference Data." Transportation **21**(2): 167-184.

Bryan, S., M. Buxton, et al. (1998). "Magnetic resonance imaging for the investigation of knee injuries: an investigation of preferences." Health Econ **7**(7): 595-603.

Carlsson, F. and P. Martinsson (2001). "Do hypothetical and actual marginal willingness to pay differ in choice experiments? Application to the valuation of the environment." Journal of Environmental Economics and Management **41**(2): 179-192.

Clark, J., J. Burgess, et al. (2000). ""I struggled with this money business": respondents' perspectives on contingent valuation." Ecological Economics **33**(1): 45-62.

DeShazo, J. R. and G. Fermo (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.

Guzman, R. M. and C. D. Kolstad (2007). "Researching preferences, valuation and hypothetical bias." Environmental & Resource Economics **37**(3): 465-487.

Gyrd-Hansen, D. and U. S. Skjoldborg (2008). "The price proxy in discrete choice experiments: Issues of relevance for future research." Chapter 8, Ryan, M., Gerard K., and Amaya-Amaya (eds.), Using Discrete Choice Experiments to Value Health and Health Care, 175-193.

Hall, J., P. Kenny, et al. (2002). "Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination." Health Econ **11**(5): 457-465.

Hanson, K., B. McPake, et al. (2005). "Preferences for hospital quality in Zambia: results from a discrete choice experiment." Health Econ **14**(7): 687-701.

Howard, K. and G. Salkeld (2009). "Does Attribute Framing in Discrete Choice Experiments Influence Willingness to Pay? Results from a Discrete Choice Experiment in Screening for Colorectal Cancer." Value in Health **12**(2): 354-363.

Jan, S., G. Mooney, et al. (2000). "The use of conjoint analysis to elicit community preferences in public health research: a case study of hospital services in South Australia." Aust N Z J Public Health **24**(1): 64-70.

Johnson, F. R. and W. H. Desvousges (1997). "Estimating stated preferences with rated-pair data: Environmental, health, and employment effects of energy programs." Journal of Environmental Economics and Management **34**(1): 79-99.

Johnson, F. R., A. F. Mohamed, et al. (2011). "How Does Cost Matter in Health-Care Discrete-Choice Experiments?" Health Econ **20**(3): 323-330.

Kjaer, T., M. Bech, et al. (2006). "Ordering effect and price sensitivity in discrete choice experiments: need we worry?" Health Econ **15**(11): 1217-1228.

Little, J. and R. Berrens (2004). "Explaining disparities between actual and hypothetical stated values: further investigation using meta-analysis " Economics Bulletin **3**: 1-13.

Lloyd, A., E. McIntosh, et al. (2005). "The importance of drug adverse effects compared with seizure control for people with epilepsy: a discrete choice experiment." Pharmacoeconomics **23**(11): 1167-1181.

Lloyd, A. J. (2003). "Threats to the estimation of benefit: are preference elicitation methods accurate?" Health Econ **12**(5): 393-402.

McIntosh, E. (2006). "Using discrete choice experiments within a cost-benefit analysis framework - Some considerations." Pharmacoeconomics **24**(9): 855-868.

McIntosh, E., C. Donaldson, et al. (1999). "Recent advances in the methods of cost-benefit analysis in healthcare - Matching the art to the science." Pharmacoeconomics **15**(4): 357-367.

McIntosh, E. and M. Ryan (2002). "Using discrete choice experiments to derive welfare estimates for the provision of elective surgery: Implications of discontinuous preferences." Journal of Economic Psychology **23**(3): 367-382.

Mozumder, P. and R. P. Berrens (2007). "Investigating hypothetical bias: induced-value tests of the referendum voting mechanism with uncertainty." Applied Economics Letters **14**(10): 705-709.

Ozdemir, S., F. R. Johnson, et al. (2009). "Hypothetical bias, cheap talk, and stated willingness to pay for health care." J Health Econ **28**(4): 894-901.

Payne, J. W., J. R. Bettman, et al. (1993). The adaptive decision maker. Cambridge ; New York, NY, USA, Cambridge University Press.

Propper, C. (1994). "The disutility of time spent on the United Kingdom's National Health Service waiting lists." Journal of Health Resources **30**(677-700).

Ratcliffe, J. (2000). "The use of conjoint analysis to elicit willingness-to-pay values. Proceed with caution?" Int J Technol Assess Health Care **16**(1): 270-275.

Roux, L., C. Ubach, et al. (2004). "Valuing the benefits of weight loss programs: an application of the discrete choice experiment." Obes Res **12**(8): 1342-1351.

Ryan, M. (2004). "Deriving welfare measures in discrete choice experiments: a comment to Lancsar and Savage (1)." Health Econ **13**(9): 909-912; discussion 919-924.

Ryan, M., J. Diack, et al. (2005). "Rapid prenatal diagnostic testing for Down syndrome only or longer wait for full karyotype: the views of pregnant women." Prenat Diagn **25**(13): 1206-1211.

Ryan, M. and K. Gerard (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. and J. Hughes (1997). "Using conjoint analysis to assess women's preferences for miscarriage management." Health Econ **6**(3): 261-273.

Ryan, M., V. Watson, et al. (2009). "Rationalising the 'irrational': a think aloud study of discrete choice experiment responses." Health Econ **18**(3): 321-336.

Ryan, M. and S. Wordsworth (2000). "Sensitivity of willingness to pay estimates to the level of attributes in discrete choice experiments." Scottish Journal of Political Economy **47**(5): 504-524.

San Miguel, F., M. Ryan, et al. (2000). "Applying conjoint analysis in Economic evaluations: an application to menorrhagia." Applied Economics **32**(823-833).

Skjoldborg, U. S. and D. Gyrd-Hansen (2003). "Conjoint analysis. The cost variable: an Achilles' heel?" Health Econ **12**(6): 479-491.

Stensrud, J., E. Sylvestre, et al. (1997). "Targeting Medicare consumers. Managed care providers can make inroads by understanding preference and cost-sensitivity issues." Mark Health Serv **17**(1): 8-17.

Szeinbach, S. L., H. L. Mason, et al. (1990). "Variables affecting pharmacists' willingness to accept third-party prescription contracts: a conjoint analysis." J Health Care Mark **10**(3): 45-50.

Taylor, S. J. and C. L. Armour (2002). "Acceptability of willingness to pay techniques to consumers." Health Expect **5**(4): 341-356.

Telser, H. and P. Zweifel (2002). "Measuring willingness-to-pay for risk reduction: an application of conjoint analysis." Health Econ **11**(2): 129-139.

van der Berg, B., M. Al, et al. (2005). "Economic valuation of informal care: The conjoint measurement method applied to informal caregiving." Social Science and Medicine **91**: 1342-1355.

van der Pol, M. and J. Cairns (1998). "Establishing patient preferences for blood transfusion support: an application of conjoint analysis." J Health Serv Res Policy **3**(2): 70-76.

Weston, A. and P. Fitzgerald (2004). "Discrete choice experiment to derive willingness to pay for methyl aminolevulinate photodynamic therapy versus simple excision surgery in basal cell carcinoma." Pharmacoeconomics **22**(18): 1195-1208.

Wooldridge, J. M. (2002). "Economic analysis of cross section and panel data." MIT press.

Chapter 4: Estimating willingness to pay using choice experiments when healthcare is free at the point of use. Are we throwing too much caution to the wind?

1. Introduction.

This chapter reports on an assessment of a further potential source of bias in the use of DCEs as a measure of WTP. The question is how the use of different descriptors for a monetary attribute might affect estimated WTP (Ratcliffe 2000). Thus, this chapter examines the impact of using different descriptors for the monetary attribute upon estimated Marginal Willingness to Pay (MWTP). Moreover, the other issue raised in the previous chapter about hypothetical bias (whether respondents take differences in the monetary attribute into account when making DCE choices) is addressed again using different DCE data.

In the interests of avoiding confusion, before outlining the theoretical background to this chapter, it is worthwhile explaining how the DCE analysis presented here, and the sample used, differs from that in Chapter 3.

In chapter 3 of the thesis, I outlined an analysis approach adopted to establish whether respondents consider the monetary attribute when they answer DCE questions. The project team (of which I was a part) had been commissioned to undertake a DCE to establish whether women with menstrual disorders would prefer to receive care via a long established Consultant-led model of care (the One-stop model) or via a new integrated care pathway which involved a selected number of GPs being trained to provide specialist care for women with menstrual disorders using a jointly agreed care pathway protocol. This new integrated model removed the traditional ‘referral’ between healthcare sectors allowing more

involvement of GPs in the provision of specialized services to women with menstrual disorders.

After undertaking the initial DCE analysis upon women with Menstrual disorders (heavy periods or menorrhagia) it was suggested to me by Dr Sophia Julian (a clinician who had interviewed the women for the pilot Menstrual disorders DCE project), that the attributes and levels we had specified for the 'Period problems' DCE questionnaire were sufficiently generic to be used to assess the preferences of women with other gynaecological complaints. Moreover, it was suggested that in this way we could readily expand our sample of DCE respondents. This would assist us in addressing the methodological problem highlighted in Chapter 3 (i.e. whether estimates of WTP might be subject to hypothetical bias because of the failure of respondents to 'factor in' changes in the monetary attribute).

To recap, for the purpose of the analysis contained in the last chapter most of the DCE respondents (n=121) had Period problems, and so filled in a questionnaire with a cover which referred to 'Health care for women with Period Problems'. However a minority of respondents (n=59), were presented with a questionnaire in which they faced the same DCE choice scenarios but, because they had a different gynaecological complaint, they instead completed a DCE questionnaire with a cover entitled 'Gynaecology health care.'

In this chapter, only DCE respondents answering the questionnaire about 'Gynaecology health care' are included in the analysis. Moreover, this follow-on DCE survey of women with other miscellaneous gynaecological problems was designed to allow us to address a further methodological question i.e. whether different ways of phrasing the monetary attribute (which do not affect who might pay) might affect estimates of WTP.

To this end I produced DCE questionnaires which were otherwise the same, except for the fact that the monetary attribute was described in 3 different ways. The 3 descriptors for the monetary attribute emerged from the DCE pilot exercise (described in Chapter 3). These were the descriptors that respondents most preferred when we asked them to indicate their preferences. Also, because this new sample of respondents might also fail to take difference in the monetary attribute into account, I again used information from a question posed to respondents about whether they took differences in the monetary attribute into account when choosing either options A or B.

However, on this occasion (in contrast to the analysis in chapter 3) because we had 3 descriptors for the monetary attribute, the question posed asked about whether differences in either 'cost to you'; 'willingness to pay'; or 'amount lost' had affected a respondent's choice of either option A or option B. Which of the 3 monetary descriptors was referred to in the question posed, was of course contingent upon how the monetary attribute was described in the early part of the questionnaire.

2. Background.

This chapter therefore reports on an assessment of two key potential sources of bias in the use of DCEs as a measure of WTP. The theoretical hypothetical bias issues surrounding our analysis about whether respondents take differences in the monetary attribute into account when making DCE choices are already extensively discussed in chapter 3 (so they are not discussed again here), although we repeat a similar analysis (using different DCE data) relating to this issue. The new issue this chapter examines which is not considered in the previous chapter however is that it is unclear how the use of different descriptors for a monetary attribute might affect estimated WTP (Ratcliffe 2000).

When applying DCEs to derive Willingness to Pay, the payment vehicle is often described in alternative ways, but descriptors usually refer to costs in some way. Examples of cost related descriptions include: 'Cost' (San Miguel, Ryan et al. 2000; Howard and Salkeld 2009; van der Pol, Shiell et al. 2009; Oteng, Marra et al. 2011; Whitty, Scuffham et al. 2011); 'Costs' (Musters, de Bekker-Grob et al. 2011); or 'Cost to you' (Ryan and Hughes 1997; Regier, Friedman et al. 2009; van der Pol, Shiell et al. 2009).

A wide range of other variations include cost-related descriptors such as: 'Total cost of MRI to the patient' (Bryan, Buxton et al. 1998); 'Monthly cost of treatment' (Darba, Restovic et al. 2011); 'Cost of treatment per year (net)' (Lloyd, Nafees et al. 2011);'Total cost to patient' (de Bekker-Grob, Essink-Bot et al. 2009); 'Payment towards the total cost of care' for antenatal care (Deverill, Lancsar et al. 2010); 'the drug will cost you an extra £…' for drug costs (Lloyd, McIntosh et al. 2005); 'Cost of the smear' (Wordsworth, Ryan et al. 2006); 'Vaccine cost for 3 doses of vaccine' (Poulos, Yang et al. 2011); 'Cost of test' for prenatal diagnosis (Chan, Sahota et al. 2009); 'Colorectal Cancer screening will cost you' (Nayaradou, Berchi et al. 2010); 'Cost of therapy programme' for post stroke rehabilitation (Laver, Ratcliffe et al. 2011); 'Personal cost to you per month not covered by insurance of rheumatoid arthritis care' (Ozdemir, Johnson et al. 2009); 'Additional cost for the hospital' (Faggioli, Scalone et al. 2011); 'Cost of medicare therapy management' (Hong, Liu et al. 2011). On other occasions indirect costs associated with the intervention, including 'travel costs' (McIntosh and Ryan 2002), have been specified as the monetary attribute.

Another popular approach for defining the monetary variable involves specifying price in some way, rather than 'cost', which is therefore an explicit way of establishing WTP. For example 'Price' (Ahmed and Fincham 2011); 'Price one has to pay for surgery' (Essers, van

Helvoort-Postulart et al. 2010); ‘Price of cigarettes’ (Goto, Takahashi et al. 2011); ‘Price: price for the complete treatment’ (Marti 2011); ‘Price of influenza vaccination’ (Pereira, Mulligan et al. 2011) ‘How much you have to pay’ (Scalone, Watson et al. 2011). Other similar descriptors include ‘The amount of money you have to spend to get the drug (£) (clinical advice provided + medicine + travel)’ (Tinelli, Ryan et al. 2009); Or alternatively, if respondents do not pay but instead might benefit, ‘Cash incentive’ for walking programmes (Brown, Finkelstein et al. 2009); or financial ‘Reward’ relating to smoking cessation (Goto, Takahashi et al. 2009)

Prices are also frequently specified in terms of out of pocket expenditure for example ‘Monthly out of pocket expenditure’ (Kjaer, Bech et al. 2006); ‘Out-of-pocket costs of medications’ (Brown, Pashos et al. 2011); ‘Out of pocket cost of treatment’(Sweeting, Whitty et al. 2011); ‘Out-of-pocket payment per month in excess of present expenditure for arthritis medication (DKK)’ (Skjoldborg, Lauridsen et al. 2009); or alternatively ‘Additional payments’ (Bogelund, Vilsboll et al. 2011); or ‘Co-payment per week’ for long term care services (Nieboer, Koolman et al. 2010).

A limited number of papers specify tax payments. Examples include ‘Increase of healthcare taxes’ (Scalone, Mantovani et al. 2009); or ‘Additional tax contribution to healthcare’ (Scuffham, Whitty et al. 2010).

The background literature indicates that there are a range of descriptors which have and can be used to describe the monetary attribute. Moreover, it is unclear to what degree the choice of descriptor can affect estimates of Willingness to Pay. There is every reason to anticipate that descriptors which, for example, refer to tax might meet with a difference response to those referring to price or cost. Therefore, estimates of WTP derived using a tax-based

monetary attribute (borne by the general public at large) may well be different from those derived using cost or price-based descriptors (for which respondents are asked to assume they bear the cost or price). What is less clear however, is whether or not other small changes in the way in which the money attribute is described (which do not affect who pays) might influence estimated WTP. The analysis presented in this chapter addresses this question.

To explore this issue, we used the 3 descriptors for the monetary attribute which had proved most popular in pilot interviews ('Cost to you'; 'Amount lost' or 'Willingness to pay') and distributed equal numbers of questionnaires with these descriptors. We then tried to establish whether estimated WTP was affected by the description of the monetary attribute.

We also tested the premise, once again that, in the context of 'free' healthcare respondents may ignore cost in their decision-making (using a similar approach to that adopted in Chapter 3). Once again, we identified the proportion of respondents who admitted they had not taken the monetary attribute into account, and then assessed the impact this had upon estimated WTP.

3. Methods.

A DCE questionnaire was produced for distribution amongst gynaecology patients attending a National Health Service hospital in Leicester, UK as part of an evaluation for the 'Bridges project.' The Bridges Project used a guideline based approach to provide an agreed integrated care pathway or bridge between primary and secondary care; this allows direct booking by GPs of investigations and treatments (i.e. surgery) for patients who remain under their care throughout (Julian, Naftalin et al. 2008). The new model therefore removes

traditional ‘referral’ between healthcare sectors, and places increased responsibility on GPs for providing specialized services.

The DCE questionnaire was designed to test the robustness of applying DCEs to elicit WTP when healthcare is free at the point of use. The DCE questionnaire was originally piloted on women with menstrual disorders, and developed in order to assess the preferences of these women for a new Bridges project model of care as compared to their preferences for the existing status quo option of Consultant-led ‘One-stop’ provision. The new integrated care GP led model (Bridges project model) was therefore compared with the existing One-Stop Consultant delivered service in secondary care, using a DCE with attributes designed to allow differences in the attributes of the models of provision to be valued in monetary terms in patients with Menstrual disorders (see chapter 3).

However because of the fact that the new integrated care GP led model (Bridges project model) could be used to provide gynaecological provision more generally (and not just for women with menstrual disorders), we also wanted to establish the preferences of gynaecological patients more generally for the two different models of provision.

Moreover, because the original DCE which was applied to women with menstrual disorders was sufficiently generic in nature, as long as the cover sheet was changed to refer to ‘Gynaecology health care’ the same DCE scenarios could be applied to patients with other miscellaneous gynaecological complaints rather than menstrual disorders patients. The results from applying the DCE upon women with a range of gynaecological complaints are therefore presented in this chapter.

3.1. Piloting and then designing the final DCE questionnaire.

The DCE questionnaire was piloted on a group of women with menstrual disorders (for details of this please refer to Section 3.2 in chapter 3). The piloting was conducted by Dr Sophia Julian. Dr Julian was at the time a junior hospital doctor (Gynaecology). In her view the final attributes and levels which were finally selected for the final DCE questionnaire (designed to be applied to patients with menstrual disorders) were sufficiently generic to be applied to women with other Gynaecological problems. This meant that a separate repeat pilot exercise was not required to develop the 'Gynaecology health care' DCE questionnaire. Instead all that was required was that the title be change to reflect the fact that the questionnaire was now being used to evaluate a range of gynaecologically related provision rather than 'Menstrual disorders.'

3.2. Attributes and levels.

The final DCE questionnaire included attributes and associated levels as described in table 1. The design was a main effects design with 16 choices generated by SPEED. Half the questionnaires had the first half of the choices generated by SPEED in, and the other half had the other half of these choices in. We used a fixed comparator, with levels for the fixed comparator set as close as possible to existing One-stop Consultant-led provision (for further details of the selection of attribute levels for the fixed comparator see section 3.3 in chapter 3). The fixed comparator assumed a 1 day wait for test results; no cost to you; care via a male consultant; a 12 week wait for an appointment; and getting to see the same doctor about half the time. After designing the choices we used SPSS to check for collinearity between attributes.

Table 1: Basic description of attributes, variable names, and levels.

Attribute	**Variable name**	**Levels**
The type of doctor	typd	GP or Consultant
The gender of the doctor	gend	Male or Female
Time waiting for an appointment to see the doctor (either GP or Consultant)	wait	1 day, 4 days, 6 weeks, 12 weeks
Continuity of care (How often you get to see the same doctor)	cont1	All of the time rather than half of the time
Continuity of care (How often you get to see the same doctor)	cont2	Half of the time rather than none of the time
How long you have to wait for test results	res	1 day, 2 days, 2 weeks, 4 weeks
The monetary attribute (either cost to you; amount lost; or willingness to pay)	mon	None, £25, £75, and £125

The attributes and levels (for everything except the monetary attribute) are identical to those specified for the 'Menstrual disorders' DCE analysis. So in the interests of avoiding repetition we refer the reader to section 3.3 of chapter 3 which provides a full explanation for the rationale of the choice of attributes and levels used in this questionnaire (other than the monetary attribute which is explained here).

The 3 descriptors for the monetary attribute emerged from the DCE pilot exercise which was described in section 3.2 of chapter 3. They were the 3 descriptors that respondents ranked most highly in a ranking exercise when we asked them to indicate which descriptor they

preferred. The 3 most preferred monetary descriptors included ‘cost to you’, ‘amount lost’, and ‘willingness to pay.’

I think it was particularly important to ensure that one of the descriptors had a descriptor which referred in some way to costs. This is because very often when a monetary attribute is used it is expressed in a manner which relates to cost. This is why the selection of ‘cost to you’ as one of the monetary descriptors was appropriate. Moreover, the descriptor ‘cost to you’ has been used before (Ryan and Hughes 1997; Regier, Friedman et al. 2009; van der Pol, Shiell et al. 2009). In addition we wanted to include a descriptor that implied a price had to be paid (because if monetary descriptors do not refer to cost they often refer to price). So the descriptor ‘willingness to pay’ was a useful one to include for that reason.

Moreover it is interesting to look at the issue of whether respondents responded differently to the descriptor ‘willingness to pay’ to ‘cost to you.’ The descriptor ‘cost to you’ (like other cost based monetary descriptors) might be open to criticism because of concerns that it might encourage cost based responses (whereby respondents value differences in terms of what they feel they should cost rather than what they think they are worth). The term ‘willingness to pay’ in contrast should encourage a response more in line with what people think differences are worth to them (i.e. thereby including the element of ‘consumer surplus’ that should be incorporated when benefits are valued). The other term ‘amount lost’ was included because it was the third most popular descriptor. Moreover, if people suppose a monetary loss, then we would assume they would require a compensating variation (improvement in other attributes) in exchange for that loss (so it is a reasonable candidate descriptor to include for the purposes of this analysis).

In chapter 3 it was explained that for the purposes of the 'Menstrual disorders' DCE we had ended up with 12 different questionnaire designs. This was because in order to avoid respondents facing too many questions we had only wanted them to face half of the 16 choices emanating from the SPEED design. Moreover, there were 6 attributes and we wanted to rotate them in the ordering of attributes within scenarios. This was because we were concerned that the ordering of attributes within scenarios might affect respondent valuation of attributes. This meant that we had 12 final DCE questionnaires for women with 'Period Problems' and the 'cost to you' attribute and attributes were rotated so that the 'cost to you' attribute was 1st in the list for 2 versions, 2nd in the list for 2 versions, 3rd in the list for 2 versions, 4th in the list for 2 versions, 5th in the list for 2 versions, and 6th for 2 versions.

In the same way for the purposes of this 'Gynaecology healthcare' DCE study there were 12 versions of the questionnaire with a 'cost to you' monetary attribute. However, because we also had questionnaires with a monetary attribute which was described as either 'amount lost' or 'willingness to pay' and because in the interests of consistency we needed 12 versions of each of these, we ended up with 36 different versions of the DCE questionnaire overall (we attempted to distribute numbers of each of the different versions as evenly as possible). As it is impractical and unnecessary to reproduce 36 questionnaires in the appendices, just 2 samples ones are contained in appendix D, one with an 'Amount Lost' monetary attribute descriptor and the other with a 'Willingness to Pay' monetary attribute descriptor. It should be noted that the DCE did not provide consistency check type data for analysis.

Once again as per the analysis in chapter 3, we did not have the intention of testing to see whether estimated WTP varied according to the positioning of the monetary attribute (because our sample may have been too small to facilitate such analysis anyway). However, we reasoned that if the positioning of the monetary attribute was varied in this consistent

manner, then estimated WTP would not be biased due to the ordering of attributes within scenarios.

3.3. More details about the final questionnaire.

Also, because respondents might also fail to take difference in the monetary attribute into account (in the same way that they did with respect to the 'Menstrual disorders DCE), as with the analysis detailed in chapter 3 we again used information from a question posed to respondents about whether they took differences in the monetary attribute into account when choosing either options A or B.

So, the questionnaire included a question to establish whether respondents factored the monetary attribute into their decision making. For those receiving the 'cost to you' variant it read:

Did differences in the amount of 'cost to you' for options A and B influence your choices? ❑ Yes ❑ No ❑ Sometimes

(Question 1)

For those receiving 'amount lost' or 'willingness to pay' variants of the questionnaire, an equivalent question was included.

3.4. Data analysis and hypothesis testing.

Data analysis was conducted using STATA in each case using Random Effects Probit. We calculated willingness to pay estimates and used the Delta method (Wooldridge 2002) which provides asymptotic confidence intervals. This was implemented using the non-linear confidence interval command in STATA v. 9.2. We conducted Wald tests using STATA.

3.4.1. Hypothesis that a basic functional form is adequate for the entire sample.

We assume a 'state of the world' model in which WTP for one alternative is to be considered (Ryan 2004). If DCE responses are not affected either by the descriptor for the monetary attribute, or the nature of patients' responses to question 1, then a simple Random Effects Probit model using the variables defined in table 1, with the following functional form (model 1) would produce estimates of WTP unaffected by these 2 potential sources of bias:

$$Y_{ij} = \alpha_0 + \alpha_1 typd_{ij} + \alpha_2 gend_{ij} + \alpha_3 wait_{ij} + \alpha_4 cont1_{ij} + \alpha_5 cont2_{ij} + \alpha_6 res_{ij} + \alpha_7 mon_{ij} + \mu_i + \xi_{ij}$$

(Model 1)

Here Y_{ij} is the binary dependent variable, from individuals i = 1…m, for observations j = 1…n_i, The number of observations n_i varies because the i individuals do not all complete every pairwise choice (a minority of respondents will not answer all choices). The term μ_i is the random effects error term (which allows for multiple responses from i respondents) and ξ_{ij} is the standard Probit error term for individuals i for j observations.

By implication because the monetary attribute is expressed in loss space (as it may be described as amount lost, cost to you, and willingness to pay), the model also assumes that a

representative and average individual may be indifferent also between detrimental changes in non-monetary attributes and monetary loss at calculated levels of Marginal Willingness to Pay.

Marginal Willingness to Pay (MWTP) can be established for the attributes h (h=1,..,6). For an individual willingness to pay (WTP_h) for a change in attribute variables h is given by:

$$MWTP_h = \alpha_h / -\alpha_7.$$

(Equation 1)

Here α_7 defines the value of a change in monetary units in loss space, therefore $-\alpha_7$ gives the value of a one unit change in monetary units in monetary gain space. For model 1 it is assumed that WTP can be established simply by establishing how much people value positive changes in the monetary attribute (mon), in indirect utility terms (which is estimated by $-\alpha_7$) and comparing this with how much people value changes in other attributes in indirect utility terms (estimated for attributes by α_h).

We therefore want to establish whether model 1 can reasonably be used to model the sample data and be accurately used to calculate MWTP. If it can it suggests that MWTP results are invariant both to the choice of descriptor for MWTP, and also to whether or not respondents factor MWTP into their decision making. The accuracy of estimates of marginal willingness to pay ($MWTP_h$) crucially depends upon obtaining a reliable estimate for α_7, and reliable estimates of coefficients on non-monetary attribute variables (α_h), which are not biased as a result of hypothetical or strategic bias.

3.4.2. Alternative hypothesis - a segmented model is required.

If it is the case that results are sensitive either to the choice of descriptors used for the monetary attribute, or how respondents respond to question 1, then an alternative model (model 2) might better represent the data. Model 2 allows for the valuation of attributes separately according to whether respondents responded to a questionnaire with either an 'amount lost' (Da), 'cost to you' (Dc), or 'willingness to pay' (Dw) descriptor for the monetary attribute. Model 2 also allows for separate calculation of MWTP amongst the sub-group of respondents who claim they did not take differences in the monetary attribute into account (Dn). Here Y_{ij} is the binary dependent variable, from individuals i = 1…m, for observations j = 1…$n_{i,}$

$$Y_{ij} = \alpha_0 Da_{ij} + \alpha_1 Da_{ij} typd_{ij} + \alpha_2 Da_{ij} gend_{ij} + \alpha_3 Da_{ij} wait_{ij} + \alpha_4 Da_{ij} cont1_{ij} + \alpha_5 Da_{ij} cont2_{ij} + \alpha_6 Da_{ij} res_{ij} + \alpha_7 Da_{ij} mon_{ij} + \alpha_8 Dc_{ij} + \alpha_9 Dc_{ij} typd_{ij} + \alpha_{10} Dc_{ij} gend_{ij} + \alpha_{11} Dc_{ij} wait_{ij} + \alpha_{12} Dc_{ij} cont1_{ij} + \alpha_{13} Dc_{ij} cont2_{ij} + \alpha_{14} Dc_{ij} res_{ij} + \alpha_{15} Dc_{ij} mon_{ij} + \alpha_{16} Dw_{ij} + \alpha_{17} Dw_{ij} typd_{ij} + + \alpha_{18} Dw_{ij} gend_{ij} + \alpha_{19} Dw_{ij} wait_{ij} + \alpha_{20} Dw_{ij} cont1_{ij} + \alpha_{21} Dw_{ij} cont2_{ij} + \alpha_{22} Dw_{ij} res_{ij} + \alpha_{23} Dw_{ij} mon_{ij} + \alpha_{24} Dn_{ij} + \alpha_{25} Dn_{ij} typd_{ij} + \alpha_{26} Dn_{ij} gend_{ij} + \alpha_{27} Dn_{ij} wait_{ij} + \alpha_{28} Dn_{ij} cont1_{ij} + \alpha_{29} Dn_{ij} cont2_{ij} + \alpha_{30} Dn_{ij} res_{ij} + \alpha_{31} Dn_{ij} mon_{ij} + \mu_i + \xi_{ij}$$

(Model 2)

The binary dependent variable Y_{ij} and the error terms μ_i, and ξ_{ij} have the same interpretation as for model 1. The model includes additive dummies (Da, Dc, Dw), to provide information upon the intercept term for each of the monetary descriptor groups respectively (amount lost, cost to you, willingness to pay), such that Da = 1 for the amount lost group, or 0 otherwise; Dc = 1 for the cost to you group, or 0 otherwise; Dw = 1 for the Willingness to pay group, or 0 otherwise. We also included (Dn) for the group that do not factor the monetary attribute into

their decision making, Dn = 1 for those who do not factor in the monetary attribute, or 0 otherwise.

A likelihood ratio test can be conducted comparing the restricted model (model 1) with the unrestricted model (model 2). If there is evidence of a jointly significant difference, then this would suggest that model 2 better represents the underlying data than model 1.

3.4.3. Establishing whether estimated Willingness to Pay is sensitive to the monetary descriptor used – in the sub-group who claim they do not take differences in the monetary attribute into account when making choices.

MWTP for each of the 3 monetary descriptor groups for the respondents who 'do not' take differences in the monetary attribute into account can be expressed (for details of variables see table 1), for m=1…6, as:

a) 'Amount lost' group: $MWTP_{m,AL} = (\alpha_m+\alpha_{m+24}) / (-\alpha_7-\alpha_{31})$

(equation 2)

b) 'Cost to you' group: $MWTP_{m,CTY}= (\alpha_{m+8}+\alpha_{m+24}) / (-\alpha_7-\alpha_{31})$

(equation 3)

c) 'Willingness to Pay' group: $MWTP_{m,WTP}= (\alpha_{m+16}+\alpha_{m+24}) / (-\alpha_7-\alpha_{31})$

(equation 4)

Wald tests 1-6 are conducted to examine the proposition that estimates of MWTP are sensitive to the choice of monetary descriptor.

The Wald test restrictions require that:

$$MWTP_{m,AL} = MWTP_{m,CTY} = MWTP_{m,WTP}$$

(equation 5)

Wald tests (1-6) which are detailed in the results section (table 8), test the restriction that estimated WTP is not sensitive to the descriptor used for the monetary attribute amongst those who 'do not' take differences in the monetary attribute into account when making pairwise choices. Evidence of a difference between the groups at the 5% significance level would require a p-value (prob > chi^2) of ≤ 0.05.

3.4.4. Establishing whether estimated Willingness to Pay is sensitive to the monetary descriptor used – for the sub-group who claim they do or 'sometimes' take differences in the monetary attribute into account when making choices.

MWTP for each of the 3 monetary descriptor groups for the respondents who do or 'sometimes' take differences in the monetary attribute into account can be expressed as:

a) 'Amount lost' group: $MWTP_{m,AL} = (\alpha_m)/ (-\alpha_7)$

(equation 6)

b) 'Cost to you' group: $MWTP_{m,CTY} = (\alpha_{m+8}) / (-\alpha_{15})$

(equation 7)

c) 'Willingness to Pay' group: $MWTP_{m,WTP} = (\alpha_{m+16}) / (-\alpha_{23})$

(equation 8)

Wald tests (for m =1…6) can also be undertaken to test the same proposition that WTP estimates are insensitive to the choice of monetary descriptor used, for the group of respondents who claim they do or 'sometimes' take differences in the monetary descriptor into account.

$$MWTP_{m,AL} = MWTP_{m,CTY} = MWTP_{m,WTP}$$

(equation 9)

The Wald test restrictions (tests 7-12) are again detailed in the results section in table 8, together with details of the restriction tested for each test using model 2. Evidence of a difference between the groups at the 5% significance level would again require a p-value (prob > chi^2) of ≤ 0.05.

3.4.5. Establishing whether results are sensitive to whether or not respondents claim not to factor the monetary attribute into their decision making, for each monetary descriptor:

We may want to establish whether or not there is a difference in estimates of MWTP comparing those who claim they 'do' or 'sometimes' take differences in MWTP into account when making pairwise choices vs. those who say they do not do so[1].

We need however to first establish whether there is a difference in estimated MWTP according to the monetary descriptor group, using the approach outlined in sections 3.4.3,

[1] As a very small number of respondents (6 / 188) failed to indicate whether they did or did not take differences in the monetary attribute into account, we subsumed their responses into the group who answered 'yes' or 'sometimes' to question 1. i.e. we did not put them in the group who answered 'no' because they had not responded with the clear answer 'no.'

and 3.4.4. If we discover there is, then we can examine the proposition that there is a difference between those who claim they 'do' or 'sometimes' take differences in MWTP into account when making pairwise choices vs. those who say they do not, separately for each of the 3 groups of questionnaires (i.e. for amount lost, cost to you, and willingness to pay descriptors). Therefore if any of the results obtained when undertaking the analyses indicated in section 3.4.3 and 3.4.4 indicate a statistically significant difference in estimated MWTP by descriptor used for the monetary attribute, then the following Wald tests (for m=1…6) should be conducted upon model 2:

a) 'Amount lost' group: $MWTP_{m,AL} = (\alpha_m) / (-\alpha_7) = (\alpha_m+\alpha_{m+24}) / (-\alpha_7-\alpha_{31})$

(equation 10)

b) 'Cost to you' group: $MWTP_{m,CTY} = (\alpha_{m+8}) / (-\alpha_{15}) = (\alpha_{m+8}+ \alpha_{m+24}) / (-\alpha_{15}-\alpha_{31})$

(equation 11)

c) 'WTP' group: $MWTP_{m,WTP} = (\alpha_{m+16}) / (-\alpha_{23}) = (\alpha_{m+16}+\alpha_{m+24}) / (-\alpha_{23}-\alpha_{31})$

(equation 12)

Once again evidence of a difference between the groups at the 5% significance level requires a p-value (prob > chi^2) of ≤ 0.05.

3.4.6. Establishing whether results are sensitive to whether or not respondents claim not to factor the monetary attribute into their decision making, using a pooled sample of all respondents:

In the event that the analyses outlined in both section 3.4.3 and section 3.4.4 do not provide any evidence that estimated MWTP varies according to the monetary descriptor used in the questionnaire, then econometric model 3 should be used:

$$Y_{ij} = \alpha_{0j} + \alpha_1 typd_{ij} + \alpha_2 gend_{ij} + \alpha_3 wait_{ij} + \alpha_4 cont1_{ij} + \alpha_5 cont2_{ij} + \alpha_6 res_{ij} + \alpha_7 mon_{ij} + \alpha_8 Dn_{ij} + \alpha_9 Dn_{ij} typd_{ij} + \alpha_{10} Dn_{ij} gend_{ij} + \alpha_{11} Dn_{ij} wait_{ij} + \alpha_{12} Dn_{ij} cont1_{ij} + \alpha_{13} Dn_{ij} cont2_{ij} + \alpha_{14} Dn_{ij} res_{ij} + \alpha_{15} Dn_{ij} mon_{ij} + \mu_i + \xi_{ij}$$

(Model 3)

Here once again the binary dependent variable Y_{ij} and the error terms μ_i, and ξ_{ij} have the same interpretation as for model 1. The model now pools all the data in such a way that it does not discriminate between respondents according to the monetary descriptor that they face. Therefore variables with the coefficients $\alpha_{0...}\alpha_7$ in model 3 are exactly the same as those in model 1, whilst those with the coefficients $\alpha_{8...}\alpha_{15}$ in model 3 are equivalent to those with the coefficients $\alpha_{24...}\alpha_{31}$ in model 2.

A likelihood ratio test can be used to test for the joint significance of the dummy variables, comparing model 1 and 3. A Wald test for joint significance of these variables can also be undertaken comparing model 1 and model 3.

Individual Wald tests (table 9) for each attribute (for h=1..6) can then be used to test the restriction that WTP does not vary between those who ‘do’ or ‘sometimes’ take differences in

the monetary attribute into account vs. those who do not take differences in the monetary attribute into account, when they make choices.

Evidence of a difference between the groups at the 5% significance level again requires a Wald test p-value (prob > chi^2) of ≤ 0.05.

4. Results.

4.1 Sample characteristics.

We distributed 218 questionnaires and obtained 188 responses (86% response rate). There were 62 completed responses from the 'cost to you' group, 61 from the 'amount lost' group, and 65 from the 'willingness to pay' group. Information on household income in the three groups is reported in table 2.

Table 2: Average household income by questionnaire type.

Type of questionnaire received	Income above Average	Income about average	Income below average	Non – responders to question
Cost to you	13 / 62 (21%)	23 / 62 (37%)	18 / 62 (29%)	8 / 62 (13%)
Amount lost	17 / 61 (28%)	12 / 61 (19.5%)	26 / 61 (42.5%)	6 / 61 (10%)
Willingness To Pay	12 / 65 (18.5%)	18 / 65 (27.5%)	25 / 65 (38.5%)	10 / 65 (15.5%)

We conducted a chi-squared test for the equality of sample distributions for the 3 monetary descriptor groups with respect to household income. The null hypothesis for this was that all

the samples for those with cost to you, amount lost, and willingness to pay monetary descriptors, had the same frequency distribution with respect to the percentage of respondents from above average income, average income, below average income, and non-response answers to the income category question. The p-value generated from that test for 6 degrees of freedom was $P \leq 0.14301$, so at the 5% level we could clearly reject the alternative hypothesis that the underlying sample distribution for the 3 monetary descriptor groups differed with respect to average household income levels according to type of monetary descriptor respondents faced. Respondents were otherwise representative of women presenting to hospital outpatient gynaecology clinics in Leicestershire with a typical cross section of gynaecological complaints referred to such departments. Detailed information about the medical conditions of respondents was not collated, because of the methodological rather than clinical focus of the study.

4.2. Responses to the question about whether the monetary attribute is considered when making choices.

Table 3 indicates the extent to which respondents reported that they factor differences in the monetary attribute into their decision making. Overall, 33.5% stated they did not factor the monetary attribute into their choices, and some small differences are apparent, contingent upon the type of monetary descriptor used.

Table 3: Whether respondents take the monetary attribute into account.

Type of questionnaire received	Differences in monetary attribute affected my choices			Non – responder to question
	Yes	No	Sometimes	
Cost to you	18 / 62 (29%)	18 / 62 (29%)	24 / 62 (39%)	2 / 62 (3%)
Amount lost	17 / 61 (28%)	24 / 61 (39%)	17 / 61 (28%)	3 / 61 (5%)
Willingness To Pay	27 / 65 (41.5%)	21 / 65 (32%)	16 / 65 (25%)	1 / 65 (1.5%)

We also conducted a chi-squared test for the equality of sample distributions for the 3 monetary descriptor groups with respect to whether or not respondents claimed to take the monetary attribute into account. The null hypothesis for this was that all the samples had the same frequency distribution with respect to the percentage of respondents who claimed 'Yes' they did take differences in the monetary attribute into account; claimed 'No' they did not take differences in the monetary attribute into account; claimed they 'Sometimes' took differences in the monetary attribute into account; or did not answer the question. The p-value generated from that test for the required 6 degrees of freedom was $P \leq 0.22035$. So at the 5% level we could clearly reject the alternative hypothesis that the underlying sample distribution differed with respect to whether or not the respondents took the monetary attribute into account.

4.3. Results (Model 1).

If we examine the results from model 1 (see table 4) in isolation without recourse to results from models 2 and 3, they look broadly reassuring. The constant is insignificant which means that there is (rather like the analysis in chapter 3) no evidence of an 'experience effect' involving a preference for the bundle of characteristics associated with the status quo option of Consultant-led 'One-stop' provision for gynaecological problems. Given the coding

ascribed to the variables, the baseline (model 1) results imply the following: respondents prefer a consultant gynaecologist to a GP for consultations (the *typd* coefficient is both positive and significant and quite large at 0.2312) and implies a WTP of £39.08 (£18.66 / £59.50) for a consultant rather than a GP; they also appear to prefer a female to a male doctor (*gend* is both positive and significant and has quite a large coefficient at 0.2226) and are WTP £37.62 (£19.63 / £55.60) for a female rather than male doctor. Avoiding an extra day spent waiting for a consultation is valued (*wait* has a significant although small coefficient of 0.072) and patients will thus pay £1.21 (£0.91 / £1.51) to avoid an extra day for an appointment to see the doctor. Avoiding waiting for test results for an extra day is also valued (*res* has a significant, larger negative coefficient of 0.0225) and patients will pay £3.79 (£2.62 / £4.97) to avoid an extra day waiting for test results. In terms of continuity, seeing the same doctor half of the time rather than none of the time is highly valued (*cont2* is significant, positive and has a large coefficient at 0.3602) and patients will pay £60.87 (£34.08 / £87.66) for it; surprisingly the difference between seeing the same doctor all of the time rather than half of the time is not significant (*cont1* has p=.353). Finally, the monetary attribute (*mon*) is significant and of the expected sign (-0.0059).

Table 4. Results - Random Effects Probit (Model 1).

Variable name / Attribute details	Coefficient	P-value	WTP	P- value
con: Constant	-0.8301[2] (-0.1883 /0.0223)[3]	0.122		0.135
typd: Type of doctor	0.2312 (0.1241 / 0.3384)	0.000	£39.08[4] (£18.66 / £59.50)[5]	0.000
gend: Gender of doctor	0.2226 (0.1107 / 0.3345)	0.000	£37.62 (£19.63 / £55.60)	0.000
wait: Wait for appointment to see doctor	0.072 (0.0057 / .0.0086)	0.000	£1.21 (£0.91 / £1.51)	0.000
cont1: Continuity of care (All rather than half time)	0.0548 (-0.0608 / 0.1704)	0.353	£9.27 (-£10.09 / £28.62)	0.348
cont2: Continuity of care (Half time not none of time)	0.3602 (0.2273 / 0.4932)	0.000	£60.87 (£34.08 / £87.66)	0.000
Res: Waiting for test results	0.0225 (0.0175 / 0.0274)	0.000	£3.79 (£2.62 / £4.97)	0.000
mon: Monetary attribute	-0.0059 (-0.0047 / -0.0071)	0.000		
Sample size: n = 188 **Log-likelihood:** -1241.8	**Mc Faddens R^2:**	0.103	**% of actual values predicted:**	68.6%

4.4. Results Model 2 and Model 3.

The results from model 2 are reported in table 5, and then estimated WTP based upon model 2 is presented in table 6 (alongside estimated WTP from model 1 so any differences in point estimates can be seen). Econometric results for model 3 and estimates of WTP for the same model are presented in table 7.

[2] Point estimate generated using Random Effects Probit in STATA v.9.2.
[3] 95% Confidence intervals for point estimates generated again using Random Effects Probit in STATA v.9.2.
[4] Willingness to Pay point estimate.
[5] Asymptotic 95% confidence intervals using the Delta method as described in Wooldridge (2002), and implemented in STATA v.9.2.

Table 5: Results - Random Effects Probit (Model 2).

Variable name	**Amount lost descriptor n = 62**		**Cost to you descriptor n=61**		**Willingness to pay descriptor n=65**		**Don't factor in monetary attribute n=63**	
	Coefficient	**P – value**	**Coefficient**	**P – value**	**Coefficient**	**P – value**	**Coefficient**	**P – value**
Type of doctor	**.0171**	**0.875**	**.2773**	**0.007**	**.1525**	**0.143**	**.2250**	**0.050***
Gender of doctor	**.2403**	**0.040***	**.1001**	**0.353**	**.3021**	**0.005***	**.0591**	**0.629**
Wait for appointment	**.2403**	**0.000***	**.0093**	**0.000***	**.0075**	**0.000***	**.0004**	**0.815**
Continuity of care 1	**-.0035**	**0.976**	**.0397**	**0.722**	**-.1937**	**0.081**	**.3225**	**0.010***
Continuity of care 2	**.4683**	**0.000***	**.2703**	**0.031***	**.5902**	**0.000***	**-.2135**	**0.137**
Waiting for test results	**.0170**	**0.001***	**.0261**	**0.000***	**.0215**	**0.000***	**.0036**	**0.491**
Monetary factor	**-.0098**	**0.000***	**-.0102**	**0.000***	**-.0093**	**0.000***	**.0093**	**0.000***
Intercept term	**-.1591**	**0.124**	**-.0204**	**0.835**	**-.1049**	**0.277**	**-.0270**	**0.810**

Mc Faddens $R^{2=}$ 0.143; Percentage of actual values accurately predicted by the model = 71.8%; Log-likelihood is -1193.5;

LR test comparison with model 1: λ = 113, Critical value for 24 degrees of freedom = 36.4

* highlights p-values which are statistically significant $\leq$ 5%

Table 6: Summary of WTP results: Marginal willingness to pay (£) from the model 2 vs. model 1.

Variable WTP descriptor	Model 2		Model 1
	Respondents who do or sometimes factor in the monetary variable	Respondents who do not factor in the monetary variable	Pooled: Whole sample
	Point estimate (95% CI)	***Point estimate (95% CI)***	***Point estimate (95% CI)***
Type of Dr (Consultant vs GP)			
Amount lost	£1.76 (-£20.13 / £23.64)	£605.53 (-£3,253.96 / £4465.02)	£39.08
Cost to you	£27.21 (£5.92 / £48.51)**	£598.46 (-£1428.81 / £2625.73)	(£18.66 / £59.50)**
Willingness to pay	£16.46 (-£6.50 / £39.43)	-£4597.15 (-£150524 / £141329.90)	
Gender of Dr (Female vs Male)			
Amount lost	£24.64 (£2.29 / £47.00)*	-£738.13 (-£5174.73 / £3,698.48)	£37.62
Cost to you	£9.83 (£30.04 / £10.39)	-£188.41 (-£799.20 / £422.38)	(£19.63 / £55.60)**
Willingness to pay	£32.61 (£11.25 / £53.97)**	£4358.07 (-£135095 / £143812)	
Waiting for appointment (per day)			
Amount lo	£0.59 (£0.28 / £0.91)**	£15.14 (-£77.70 / £107.97)	
Cost to you	£0.92 (£0.61 / £1.22)**	£11.48 (-£25.72 / £48.68)	-£1.21
Willingness to pay	£0.82 (£0.51 / £1.12)**	-£95.58 (-£3,142.51 / £2,942.51)	(-£0.91 / -£1.51)**
Continuity of care (all vs ½ the time)			
Amount lost			
Cost to you	-£0.36 (-£23.55 / £22.83)	£786.45 (-£4043.40 / £5616.29)	£9.27
Willingness to pay	£3.90 (-£17.49 / £25.28)	£428.53 (-£981.75 / £1838.83)	(-£10.09 / £28.62)
	-£20.92 (-£45.55 / £3.73)	-£1553.57 (-£51331 / £48224)	
Continuity of Care (1/2 vs nil time)			
Amount lost	£48.02 (£17.74 / £78.30)**	£628.01 (-£3432.99 / £4689.02)	£60.87
Cost to you	£26.53 (£0.74 / £52.31)*	£67.21 (-£376.17 / £510.58)	(£34.08 / £87.66)**
Willingness to pay	£63.72 (£30.89 / £96.54)**	-£4544.85 (-£148,791 / £139,702)	
Wait test results (per day)			
Amount lost	£1.74 (£0.61 / £2.88)**	£50.81 (£-268.76 / £370.37)	£3.79
Cost to you	£2.56 (£1.46 / £3.66)**	£35.19(-£82.98 / £153.37)	(£2.62 / £4.97)**
Willingness to pay	£2.32 (£1.17 / £3.47)**	-£303.10 (-£9922.44 / £9,316.25)	

Confidence intervals (CI) are in brackets after point estimates for WTP. They are 95% CIs derived using the Delta method (Wooldridge 2002).

Table 7: Results - Random Effects Probit (Model 3).

Variable name	Those who say they 'do' or 'sometimes' take the monetary attribute into account (n = 125)		MWTP for those who say they do or 'sometimes' take the monetary attribute into account (n= 126)		Those who say they don't factor in the monetary attribute (n=63)		MWTP for those who say they don't factor in the monetary attribute (n=63)	
	Coefficient	**P – value**	**MWTP**	**P – value**	**Coefficient**	**P – value**	**MWTP**	**P – value**
Type of doctor	**.1494**	**0.031***	**£15.62 (£0.91 / 30.32)**	**0.037***	**.2203**	**0.057**	**£1325.60 (-£8028 / £10679)**	**0.781**
Gender of doctor	**.2073**	**0.004****	**21.68 (£7.46 / 35.89)**	**0.003****	**.0667**	**0.582**	**£982.54 (-£5723 / £7688)**	**0.774**
Wait for appointment	**.0075**	**0.000****	**0.79 (£0.58 / 0.99)**	**0.000****	**.0002**	**0.918**	**£27.55 (-£163 / £218)**	**0.778**
Continuity of care 1	**-.0530**	**0.474**	**-5.55 (-£20.88 / £9.78)**	**0.478**	**.3133**	**0.012***	**£933.45 (-£5545 /-£7412)**	**0.778**
Continuity of care 2	**.4442**	**0.000****	**46.46 (£26.75 / £66.16)**	**0.000****	**-.2104**	**0.140**	**£838.45 (-£5182 / £6859)**	**0.785**
Waiting for test results	**.0218**	**0.000****	**£2.28 (£1.51 / £3.04)**	**0.000****	**.0030**	**0.562**	**£89.10 (-£717 / £539)**	**0.781**
Monetary factor	**-.0096**	**0.000****	**1**		**1**			
Intercept term	**-.0912**	**0.161**			**-.0302**	**0.788**		

Mc Faddens R^2 = 0.138; Percentage of actual values accurately predicted by the model = 71.7%; Log-likelihood is 1201.04;
LR test comparison model 3 vs. model 2: λ = 15.1, Critical value for 16 degrees of freedom = 26.3; LR test comparing model 3 vs. model 1: λ = 97.9, Critical value for 8 degrees of freedom = 15.51; Wald test comparing model 3 vs. model 2: p=0.0000;* highlights p-values which are statistically significant ≤ 5% but not at the 1% level; ** highlights p-values which are statistically significant

4.5. Comparison of model 1 vs. model 2; model 1 vs. model 3; and model 2 vs. model 3 using a likelihood ratio (LR) test, models 1 vs. 3 using a Wald test and measures of 'goodness of fit'.

We calculated the likelihood ratios (LR) for model 1 (the restricted model) and model 2 (the unrestricted model). These figures are -1250 and -1193.5 respectively. This suggests a value for λ of 113. This compares with the appropriate critical value for 24 degrees of freedom of 36.4. We can therefore reject the null hypothesis that the restricted model (model 1) provides an adequate characterization of the underlying data on the basis of the LR test.

Comparing likelihood ratios to obtain λ in order to compare model 1 (restricted model) vs. model 3 (the unrestricted model) we have figures of -1250 and -1201.04 respectively. This suggests a value for λ of 97.9. This compares with the appropriate critical value for 8 degrees of freedom of 15.5. We can therefore once again reject the null hypothesis that the restricted model provides an adequate characterization of the underlying data on the basis of the LR test. This finding was confirmed by an equivalent Wald test, which tested the hypothesis that all the dummy variables = 0, the hypothesis was very clearly rejected (p=0.0000).

The findings from the comparison of model 1 vs. model 3 (using both LR and Wald tests for joint significance) therefore provide evidence that there is a jointly significant difference in coefficients which can be attributed to a difference between those who 'do' or 'sometime' take differences in estimated WTP into account.

Using an LR test to compare model 3 (restricted model) vs. model 2 (the unrestricted model) we have figures of 1201.04 and 1193.49 respectively. This suggests a value for λ of 15.1. This compares with the appropriate critical value for 16 degrees of freedom of 26.3. We

therefore cannot reject the null hypothesis that the restricted model provides an adequate characterization of the underlying data on the basis of the LR test. This implies that estimated WTP may not be sensitive to the choice of monetary descriptor used.

Findings from LR tests (model 1 vs. model 2; and model 1 vs. model 3; and model 3 vs. model 2) therefore suggest that allowing for a difference in the econometric model between those who say they do or sometimes take differences in the monetary attribute into account vs. those who don't, is worthwhile (model 3 rather than model 1). However, adopting a model which also allows for differences in estimated WTP according to monetary descriptor (model 2) rather than just according to whether people claim they 'do' or 'sometimes' take differences in the monetary attribute into account vs. those who don't (model 3) is not justified based upon tests for joint significance (the dummy variables in model 2 which are not in model 3 are not jointly significant).

4.6. Marginal Willingness to Pay results.

Table 4, provides estimates of MWTP for model 1, alongside regression results for model 1. Table 5, details the econometric results obtained when estimating model 2, whilst table 6 summarises the willingness to pay results for models 1 and 2, so that differences in point estimates for MWTP and the significance or otherwise of variables are apparent comparing the 2 models. The 95% confidence intervals upon point estimates of MRS, are derived using the Delta method (Wooldridge 2002). Point estimates for MWTP which are significant at the 1% level are denoted by 2 asterisks, and those significant at the 5% but not 1% level are denoted by 1 asterisk. Table 6 therefore provides details of MWTP results from the pooled model (model 1), and MWTP results obtained when sub-dividing sample into the 3 monetary descriptors used i.e. 'amount lost', 'cost to you' and 'willingness to pay', as well as between

those who say they do or sometimes take differences in the monetary attribute into account vs. they do not take differences in the monetary attribute into account (model 2). The most striking finding is that none of the point estimates of MWTP in model 2 are significantly different from zero at the 5% significance level for respondents who claim not to take differences in the monetary attribute into account when making choices. This is what we might expect if these respondents are unwilling to register a willingness to pay.

In table 7, we present the results for model 3. We find that whilst 5 / 6 of point estimates of MWTP are statistically significantly different from zero at the 5% level for the group who 'do' or 'sometimes' take the monetary attribute into account when making choices, only 1 / 6 of these point estimates for MWTP are significant for the group who say they don't factor differences in the monetary attribute into account when making choices.

4.7. Use of Wald tests to establish whether estimated MWTP for attributes is related to the monetary descriptor used.

Using Wald tests we can establish whether there is evidence that estimated MWTP for attributes is related to the choice of monetary descriptor used conducting the tests outlined in section 3.4.3 and 3.4.4. We already know from a comparison of model 2 versus model 3, that tests for joint significance do not support the proposition that distinguishing by monetary descriptor group (model 2) rather than not doing this (model 3), results in a jointly significant difference.

Results of the Wald tests are detailed in table 8 below. Differences which are significant at the 1% level are denoted by 2 asterisks, and those significant at the 5% but not 1% level are

denoted by 1 asterisk. The absence of asterisks indicates that any differences which may exist are not statistically significant.

Table 8: Wald tests. – Test hypothesis results insensitive to monetary descriptor used for the group who 'do not' factor differences in the monetary attribute into their decision making (tests 1-6), and those who do (tests 7-12).

Wald test	Restriction tested (Model 2)	Results: Amount lost n= 61 Cost to you n= 62 WTP n = 65.	Wald test	Restriction tested (Model 2)	Results: Amount lost n= 61 Cost to you n= 62 WTP n = 65.
1	$(\alpha_1+\alpha_{25}) / (-\alpha_7-\alpha_{31})=$ $(\alpha_9+\alpha_{25}) / (-\alpha_{15}-\alpha_{31})=$ $(\alpha_{17}+\alpha_{25}) / (-\alpha_{23}-\alpha_{31})$	p = 0.2187	7	$(\alpha_1) / (-\alpha_7)=$ $(\alpha_9) / (-\alpha_{15})=$ $(\alpha_{17}) / (-\alpha_{23})$	p = 0.2057
2	$(\alpha_2+\alpha_{26}) / (-\alpha_7-\alpha_{31}) =$ $(\alpha_{10}+\alpha_{26}) / (-\alpha_{15}-\alpha_{31}) =$ $(\alpha_{18}+\alpha_{26}) / (-\alpha_{23}-\alpha_{31})$	p = 0.2835	8	$(\alpha_2) / (-\alpha_7) =$ $(\alpha_{10}) / (-\alpha_{15}) =$ $(\alpha_{18}) / (-\alpha_{23})$	p = 0.2536
3	$(\alpha_3+\alpha_{27}) / (-\alpha_7-\alpha_{31}) =$ $(\alpha_{11}+\alpha_{27}) / (-\alpha_{15}-\alpha_{31}) =$ $(\alpha_{19}+\alpha_{27}) / (-\alpha_{23}-\alpha_{31})$	p = 0.2001	9	$(\alpha_3) / (-\alpha_7) =$ $(\alpha_{11}) / (-\alpha_{15}) =$ $(\alpha_{19}) / (-\alpha_{23})$	p = 0.2611
4	$(\alpha_4+\alpha_{28}) / (-\alpha_7-\alpha_{31}) =$ $(\alpha_{12}+\alpha_{28}) / (-\alpha_{15}-\alpha_{31}) =$ $(\alpha_{20}+\alpha_{28}) / (-\alpha_{23}-\alpha_{31})$	p = 0.2659	10	$(\alpha_4) / (-\alpha_7) =$ $(\alpha_{12}) / (-\alpha_{15}) =$ $(\alpha_{20}) / (-\alpha_{23})$	p = 0.2397
5	$(\alpha_5+\alpha_{29}) / (-\alpha_7-\alpha_{31}) =$ $(\alpha_{13}+\alpha_{29}) / (-\alpha_{15}-\alpha_{31}) =$ $(\alpha_{21}+\alpha_{29}) / (-\alpha_{23}-\alpha_{31})$	p = 0.1566	11	$(\alpha_5) / (-\alpha_7) =$ $(\alpha_{13}) / (-\alpha_{15}) =$ $(\alpha_{21}) / (-\alpha_{23})$	p = 0.1599
6	$(\alpha_6+\alpha_{30}) / (-\alpha_7-\alpha_{31}) =$ $(\alpha_{14}+\alpha_{30}) / (-\alpha_{15}-\alpha_{31}) =$ $(\alpha_{22}+\alpha_{30}) / (-\alpha_{23}-\alpha_{31})$	p = 0.4416	12	$(\alpha_6) / (-\alpha_7) =$ $(\alpha_{14}) / (-\alpha_{15}) =$ $(\alpha_{22}) / (-\alpha_{23})$	p = 0.5196

It can be clearly seen that for all 12 Wald tests, we cannot reject any of the restrictions at the 5% significance level. There is therefore no evidence of a significant difference in MWTP for any attribute comparing different descriptors used for the monetary attribute (none of the p-values ≤ 0.05). This finding holds both amongst respondents who claim they do not take differences in the monetary attribute into account (tests 1-6), and amongst those who say they 'do' or 'sometimes' take differences in the monetary attribute into account (tests 7-12). This further reinforces the findings obtained using likelihood ratio tests.

Since there is no evidence that MWTP varies according to the descriptor used for the monetary attribute, it is appropriate to proceed to use model 3 to establish whether estimates of WTP varies comparing those who 'do' or 'sometimes' take differences in the monetary attribute into account vs. those who do not. This means that conducting the Wald tests outlined in section 3.4.5 on model 2 is unnecessary. The results (tests 13-18) are presented in table 9 below. Differences which are significant at the 1% level are denoted by 2 asterisks, and those significant at the 5% but not 1% level are denoted by 1 asterisk.

Table 9: Wald tests – Are estimates of MWTP biased due to inclusion of respondents that say they 'do not' take differences in the monetary attribute into account ('amount lost' and 'cost to you' respondents).

Wald Test	'WTP' group - Restriction tested (Model 3)	Results n = 65	Wald Test	Whole sample – Restriction tested (Model 3)	Results n = 188
13	$(\alpha_1) / (-\alpha_7) = (\alpha_1+\alpha_9) / (-\alpha_7-\alpha_{15})$	p = 0.6479	16	$(\alpha_4) / (-\alpha_7) = (\alpha_4+\alpha_{12}) / (-\alpha_7-\alpha_{15})$	p = 0.0487*
14	$(\alpha_2) / (-\alpha_7) = (\alpha_2+\alpha_{10}) / (-\alpha_7-\alpha_{15})$	p = 0.4381	17	$(\alpha_5) / (-\alpha_7) = (\alpha_5+\alpha_{13}) / (-\alpha_7-\alpha_{15})$	p = 0.0026*
15	$(\alpha_3) / (-\alpha_7) = (\alpha_3+\alpha_{11}) / (-\alpha_7-\alpha_{15})$	p = 0.0020**	18	$(\alpha_6) / (-\alpha_7) = (\alpha_6+\alpha_{14}) / (-\alpha_7-\alpha_{15})$	p = 0.0183*

Wald tests for equality of MWTP comparing those who 'do' or 'sometimes' take differences in MWTP into account relative to those who say they do not (tests 13 – 18) provide evidence that estimated MWTP varies between the groups . Although tests 13 and 14 do not demonstrate a clear statistically significant difference in estimated MWTP between the two groups, for the attributes typd and gend, tests 15-18 all suggest a statistically significant difference at the 5% level for the remaining attributes. Therefore these findings from Wald tests, confirm the findings from tests of joint significance (LR and Wald tests comparing model 3 with model 1), and suggest that estimated MWTP does vary between those who 'do' or 'sometimes' take differences in MWTP into account vs. those who do not, with respect to 4 / 6 attributes at the 5% significance level.

5. Discussion.

Firstly, our findings provide no evidence of statistically significant differences in estimates of MWTP by type of monetary descriptor using Wald tests for individual variables (table 8), or using tests for joint significance There is every reason to suppose that preferences and also MWTP should be sensitive to the issue of who bears a cost (Bryan and Dolan 2004). Consequently, we would anticipate that if the descriptor changed in such a way that it affected who paid, or the method of payment (i.e. tax vs. out of pocket expenditure), then differences in estimated MWTP between groups (according to monetary descriptor) would be apparent.

Moreover, in section 3.2 of this chapter I speculated that the use of a cost based monetary descriptor such as 'cost to you' might be more likely to encourage 'cost based responses' (whereby respondents try to value differences in 'cost to you' according to what they think improvements in other attributes might cost). In contrast, if the descriptor 'willingness to pay'

was used, I speculated it would be more likely to encourage a response which related to how much respondents value things (i.e. thereby including the element 'consumer surplus' which cost based responses might exclude). If these two premises hold, I would expect to find evidence that estimated MWTP obtained using a 'willingness to pay' descriptor might be higher than estimates obtained using a 'cost to you' descriptor. Therefore, the finding that there was no evidence that estimated MWTP varied according to choice of monetary descriptor used is reassuring. This is because it may mean that some descriptors which refer to 'cost' can be used without them necessarily encouraging 'cost based responses.' Since 'cost based responses' strip out 'consumer surplus' from benefits estimates, we would not want to conduct MWTP analysis using information from 'cost based responses.' The observed results mean that the use of a monetary descriptor that refers to cost or costs will not necessarily introduce bias. Moreover, when applying DCEs where healthcare is free at the point of use, it is common to use another descriptor other than willingness to pay for the monetary attribute. The results here suggest that, so long as this does not impact upon who pays, this need not significantly affect estimated MWTP results. That said the results here are based upon a relatively small sample. Ideally therefore we would want to conduct a similar analysis using a larger sample to be fully assured that our results would still hold.

Secondly, rather like the results of the previous analysis discussed in chapter 3, we found that about a third of respondents (33.5%) claimed that they had not factored differences in the monetary attribute into their decision making. This occurred irrespective of which descriptor was used (i.e. 29% for 'cost to you'; 39% for amount lost; and 32% for 'willingness to pay'). Of course, it could be argued that the existence of a large proportion of respondents who do not take differences in the monetary attribute into account when making decisions could be a by-product of inappropriately pitching the levels for the monetary attribute. I fully concede that this may be the case. Moreover, if I were to conduct a similar analysis again I would use the

payment card method during the pilot exercise to try to establish the most appropriate range for the monetary attribute for the final questionnaire. This might reduce the number of respondents failing to consider the monetary attribute in their decision making. However, even if I did this I could not be sure that the price vector we assumed for the monetary attribute would necessarily be appropriate because it is unclear that a definitive 'correct' vector can be established, especially for services which are not traded. Note also that inappropriate price vectors lead to hypothetical bias defined in terms of the definition we used (i.e. hypothetical bias as the error arising when questions do not elicit responses consistent with actual behaviour, because of the hypothetical nature of questions (Champ and Bishop 2001). An important point to make is that, just because some of the respondents who said they did not take differences in the monetary attribute into account may have taken these differences into account if the range for the monetary attribute had been set differently (following the use of a pilot payment card exercise), some hypothetical bias may remain if the payment vector is not appropriate for every respondent. What it does mean though is that the extent of hypothetical bias might be lower if more robust methods were to be used to set the range for the monetary attribute than those deployed here. Moreover, if we could arrive at a WTP question which was more incentive compatible than the one used here, this might reduce the scale of hypothetical bias. However, it is more difficult to develop incentive compatible WTP questions to value healthcare when it is free at the point of use.

Thirdly, we demonstrated using Wald tests in Model 3 (table 9) that the value of coefficients used to calculate WTP varied significantly between individuals who do or 'sometimes' factor in the monetary attribute vs. those who do not factor in the monetary attribute, for 4 / 6 attributes. Therefore, the inclusion of a significant minority of respondents who fail to take into account differences in the monetary attribute may bias results in the way it has done in this WTP analysis. It would therefore be interesting to establish in future analyses whether similar

results emerge following more elaborate and robust attempts to establish a sensible range for monetary attributes (such as the use of payment cards etc) and also in the presence or absence of 'cheap talk' (Ozdemir, Johnson et al. 2009).

Moreover, if I repeated this analysis again, I would establish econometrically (using an appropriate methodology (Gyrd-Hansen and Skjoldborg 2008)) whether respondents responded to the presence or absence of a non-zero price proxy, or to differences in the levels of the price proxy, or both. There is also evidence to suggest that the cost gradient may not be continuous and linear (Johnson, Ozdemir et al. 2010) as is commonly supposed when attempting to estimate WTP. This suggests the need to explore alternative specifications for estimating cost when calculating WTP. It has been suggested (Johnson, Ozdemir et al. 2010) that respondents may deploy heuristics which involve recoding costs into categories such as low, medium, and high, thereby undermining the validity of WTP estimates. If this is the case it is argued incorporating 'cheap talk' or graphic representations of attribute cost levels in questionnaire preambles might prevent this (Johnson, Ozdemir et al. 2010). Indeed it has been shown (Ozdemir, Johnson et al. 2009) that the cost function for a cheap talk subsample appears to be linear in contrast to the cost function for the main sample of respondents not engaged in 'cheap talk'. It also follows that if respondents do deploy recoding heuristics then different price vectors may trigger different recoding heuristics, underlining the importance of rigorous attempts to establish an appropriate price vector for the monetary attribute.

However, as discussed in Chapter 3, the policy implications of our findings are not clear cut. The reason for this is that incentive compatibility of the monetary attributes used for this analysis is open to question. Indeed, if I were to repeat a similar analysis again I would make sure that the questionnaire preamble did not ask respondents to "Please assume you would

not lose this amount of money even if you would not", and I would pay particular attention to ensuring that I framed and selected monetary attributes in a way that ensured that they were as 'incentive compatible' as possible. In addition a strong qualitative agenda (alongside use of the approach suggested here, to establish whether respondents take differences in the monetary attribute into account) would be helpful. This could involve the use of 'think aloud' experiments being undertaken as respondents fill in DCE questionnaires. Such exercises might shed light upon whether respondents who indicated either they did or did not take differences in the monetary attribute into account were usually being truthful or not.

6. Conclusions.

The broad conclusion of the analysis presented in this chapter is that describing a monetary attribute differently in a manner which does not impact upon the issue of 'who pays' will not necessarily affect estimates of WTP. Reassuringly, the descriptor which referred to costs ('cost to you') was associated with similar estimated levels of WTP to those obtained using the other 2 descriptors (including one which referred to 'willingness to pay'). This suggests that using a monetary descriptor which refers to cost or some variation upon cost (such as 'cost to you') does not necessarily encourage 'cost based responses.'

Considering the findings of both Chapter 3 and Chapter 4 suggests that a substantive minority (about a third) of respondents claim not to take the monetary attribute into account when making DCE choices. Moreover, estimated WTP for this group differs when compared to respondents who do or sometimes take differences in the monetary attribute into account. These results are likely to be partly a by-product of a lack of incentive compatibility in the descriptions surrounding the monetary attribute (i.e. respondents are told to "Please assume you would lose this amount of money even if you would not" in questionnaire preambles). So

there may be a case for repeating the analysis again using either a more incentive compatible monetary attribute (or if the aim is to again assess differences according to choice of monetary descriptor, more incentive compatible monetary attributes). Such an analysis though would need to address the issue of how the levels for price vector specified for the monetary attributes might affect whether respondents take the monetary attribute into account. It should also be accompanied by 'think aloud' exercises and qualitative interviews which can probe why respondents do or do not take differences in the monetary attribute into account when making DCE choices.

7. References.

Ahmed, A. and J. E. Fincham (2011). "Patients' view of retail clinics as a source of primary care: boon for nurse practitioners?" J Am Acad Nurse Pract **23**(4): 193-199.

Bogelund, M., T. Vilsboll, et al. (2011). "Patient preferences for diabetes management among people with type 2 diabetes in Denmark - a discrete choice experiment." Curr Med Res Opin **27**(11): 2175-2183.

Brown, D. S., E. A. Finkelstein, et al. (2009). "Estimating older adults' preferences for walking programs via conjoint analysis." Am J Prev Med **36**(3): 201-207 e204.

Brown, T. M., C. L. Pashos, et al. (2011). "The perspective of patients with haemophilia with inhibitors and their care givers: preferences for treatment characteristics." Haemophilia **17**(3): 476-482.

Bryan, S., M. Buxton, et al. (1998). "Magnetic resonance imaging for the investigation of knee injuries: an investigation of preferences." Health Econ **7**(7): 595-603.

Bryan, S. and P. Dolan (2004). "Discrete choice experiments in health economics. For better or for worse?" Eur J Health Econ **5**(3): 199-202.

Champ, P. A. and R. C. Bishop (2001). "Donation payment mechanisms and contingent valuation: An empirical study of hypothetical bias." Environmental & Resource Economics **19**(4): 383-402.

Chan, Y. M., D. S. Sahota, et al. (2009). "Chinese women's preferences for prenatal diagnostic procedure and their willingness to trade between procedures." Prenat Diagn **29**(13): 1270-1276.

Darba, J., G. Restovic, et al. (2011). "Patient preferences for osteoporosis in Spain: a discrete choice experiment." Osteoporos Int **22**(6): 1947-1954.

de Bekker-Grob, E. W., M. L. Essink-Bot, et al. (2009). "Preferences of GPs and patients for preventive osteoporosis drug treatment: a discrete-choice experiment." Pharmacoeconomics **27**(3): 211-219.

Deverill, M., E. Lancsar, et al. (2010). "Antenatal care for first time mothers: a discrete choice experiment of women's views on alternative packages of care." Eur J Obstet Gynecol Reprod Biol **151**(1): 33-37.

Essers, B. A. B., D. van Helvoort-Postulart, et al. (2010). "Does the Inclusion of a Cost Attribute Result in Different Preferences for the Surgical Treatment of Primary Basal Cell Carcinoma? A Comparison of Two Discrete-Choice Experiments." Pharmacoeconomics **28**(6): 507-520.

Faggioli, G., L. Scalone, et al. (2011). "Preferences of patients, their family caregivers and vascular surgeons in the choice of abdominal aortic aneurysms treatment options: the PREFER study." Eur J Vasc Endovasc Surg **42**(1): 26-34.

Goto, R., Y. Takahashi, et al. (2011). "Changes in Smokers' Attitudes Toward Intended Cessation Attempts in Japan." Value in Health **14**(5): 785-791.

Goto, R., Y. Takahashi, et al. (2009). "A cohort study to examine whether time and risk preference is related to smoking cessation success." Addiction **104**(6): 1018-1024.

Gyrd-Hansen, D. and U. S. Skjoldborg (2008). "The price proxy in discrete choice experiments: Issues of relevance for future research." Chapter 8, Ryan, M., Gerard K., and Amaya-Amaya (eds.), Using Discrete Choice Experiments to Value Health and Health Care, 175-193.

Hong, S. H., J. Liu, et al. (2011). "Conjoint analysis of patient preferences on Medicare medication therapy management." Journal of the American Pharmacists Association **51**(3): 378-387.

Howard, K. and G. Salkeld (2009). "Does Attribute Framing in Discrete Choice Experiments Influence Willingness to Pay? Results from a Discrete Choice Experiment in Screening for Colorectal Cancer." Value in Health **12**(2): 354-363.

Johnson, F. R., S. Ozdemir, et al. (2010). "Effects of simplifying choice tasks on estimates of taste heterogeneity in stated-choice surveys." Soc Sci Med **70**(2): 183-190.

Julian, S., N. J. Naftalin, et al. (2008). "An integrated care pathway for menorrhagia across the primary-secondary interface: patients' experience, clinical outcomes, and service utilization." Quality and Safety in Health Care **16**: 110-115.

Kjaer, T., M. Bech, et al. (2006). "Ordering effect and price sensitivity in discrete choice experiments: need we worry?" Health Econ **15**(11): 1217-1228.

Laver, K., J. Ratcliffe, et al. (2011). "Early Rehabilitation Management after Stroke: What Do Stroke Patients Prefer?" Journal of Rehabilitation Medicine **43**(4): 354-358.

Lloyd, A., E. McIntosh, et al. (2005). "The importance of drug adverse effects compared with seizure control for people with epilepsy: a discrete choice experiment." Pharmacoeconomics **23**(11): 1167-1181.

Lloyd, A., B. Nafees, et al. (2011). "Willingness to pay for improvements in chronic long-acting insulin therapy in individuals with type 1 or type 2 diabetes mellitus." Clin Ther **33**(9): 1258-1267.

Marti, J. (2011). "Assessing preferences for improved smoking cessation medications: a discrete choice experiment." Eur J Health Econ.

McIntosh, E. and M. Ryan (2002). "Using discrete choice experiments to derive welfare estimates for the provision of elective surgery: Implications of discontinuous preferences." Journal of Economic Psychology **23**(3): 367-382.

Musters, A. M., E. W. de Bekker-Grob, et al. (2011). "Women's perspectives regarding subcutaneous injections, costs and live birth rates in IVF." Hum Reprod **26**(9): 2425-2431.

Nayaradou, M., C. Berchi, et al. (2010). "Eliciting Population Preferences for Mass Colorectal Cancer Screening Organization." Medical Decision Making **30**(2): 224-233.

Nieboer, A. P., X. Koolman, et al. (2010). "Preferences for long-term care services: Willingness to pay estimates derived from a discrete choice experiment." Social Science & Medicine **70**(9): 1317-1325.

Oteng, B., F. Marra, et al. (2011). "Evaluating societal preferences for human papillomavirus vaccine and cervical smear test screening programme." Sex Transm Infect **87**(1): 52-57.

Ozdemir, S., F. R. Johnson, et al. (2009). "Hypothetical bias, cheap talk, and stated willingness to pay for health care." J Health Econ **28**(4): 894-901.

Pereira, C. C., M. Mulligan, et al. (2011). "Determinants of influenza vaccine purchasing decision in the US: a conjoint analysis." Vaccine **29**(7): 1443-1447.

Poulos, C., J. C. Yang, et al. (2011). "Mothers' preferences and willingness to pay for HPV vaccines in Vinh Long Province, Vietnam." Social Science & Medicine **73**(2): 226-234.

Ratcliffe, J. (2000). "The use of conjoint analysis to elicit willingness-to-pay values. Proceed with caution?" Int J Technol Assess Health Care **16**(1): 270-275.

Regier, D. A., J. M. Friedman, et al. (2009). "Valuing the benefit of diagnostic testing for genetic causes of idiopathic developmental disability: willingness to pay from families of affected children." Clin Genet **75**(6): 514-521.

Ryan, M. (2004). "Deriving welfare measures in discrete choice experiments: a comment to Lancsar and Savage (1)." Health Econ **13**(9): 909-912; discussion 919-924.

Ryan, M. and J. Hughes (1997). "Using conjoint analysis to assess women's preferences for miscarriage management." Health Econ **6**(3): 261-273.

San Miguel, F., M. Ryan, et al. (2000). "Applying conjoint analysis in Economic evaluations: an application to menorrhagia." Applied Economics **32**(823-833).

Scalone, L., L. G. Mantovani, et al. (2009). "Patients', physicians', and pharmacists' preferences towards coagulation factor concentrates to treat haemophilia with inhibitors: results from the COHIBA Study." Haemophilia **15**(2): 473-486.

Scalone, L., V. Watson, et al. (2011). "Evaluation of patients' preferences for genital herpes treatment." Sexually Transmitted Diseases **38**(9): 802-807.

Scuffham, P. A., J. A. Whitty, et al. (2010). "Health system choice: a pilot discrete-choice experiment eliciting the preferences of British and Australian citizens." Appl Health Econ Health Policy **8**(2): 89-97.

Skjoldborg, U. S., J. Lauridsen, et al. (2009). "Reliability of the Discrete Choice Experiment at the Input and Output Level in Patients with Rheumatoid Arthritis." Value in Health **12**(1): 153-158.

Sweeting, K. R., J. A. Whitty, et al. (2011). "Patient preferences for treatment of achilles tendon pain: results from a discrete-choice experiment." Patient **4**(1): 45-54.

Tinelli, M., M. Ryan, et al. (2009). "Patients' preferences for an increased pharmacist role in the management of drug therapy." Int J Pharm Pract **17**(5): 275-282.

van der Pol, M., A. Shiell, et al. (2009). "Eliciting individual preferences for health care: a case study of perinatal care." Health Expect **13**(1): 4-12.

Whitty, J. A., P. A. Scuffham, et al. (2011). "Public and decision maker stated preferences for pharmaceutical subsidy decisions: a pilot study." Appl Health Econ Health Policy **9**(2): 73-79.

Wooldridge, J. M. (2002). "Economic analysis of cross section and panel data." MIT press.

Wordsworth, S., M. Ryan, et al. (2006). "Women's preferences for cervical cancer screening: a study using a discrete choice experiment." Int J Technol Assess Health Care **22**(3): 344-350.

Chapter 5: Who should be prioritized for renal transplantation?: Analysis of key stakeholder preferences using discrete choice experiments.

1. Introduction.

I was the principal investigator for the UK's first DCE study related to establishing allocation criteria for kidney transplantation. This chapter both summarizes the methodology applied, as well as some of the results and conclusions emanating from this study.

In addition to collecting data on the DCE preferences of renal patients, the project team also wanted to gather data relating to the preferences of other key stakeholders including healthcare professionals; live kidney donors / relatives of deceased kidney donors; and those who care for renal transplant patients. Moreover, ethnic minority patients are more susceptible to many diseases that lead to renal disease, and organ donation rates are also lower amongst ethnic minority groups. So if ethnic minority patients require a kidney transplant they may be less likely to get one from a closely matching donor. Therefore, we needed to assess whether the preferences of ethnic minority patients differed from those of other patients.

We also wanted to be able to establish whether the preferences of respondents of healthcare professionals, carers and donor groups differed from patients. This was in the same way that we needed to establish whether ethnic minority patients had different preferences to other patients. To do this rigorously, we needed to establish whether any differences in the way in which non-patient respondent groups valued a particular attribute relative to the patient group, were statistically significant. Also we needed to establish whether ethnic minorities had

statistically significantly different preferences for attributes compared to non-ethnic minority patients.

At a methodological level (as indicated in Chapter 1, section E1.3) there are therefore numerous issues related to preference heterogeneity that needed to be addressed for the purposes of this research. In Chapter 1 (section E1.3) I pointed out that a variety of approaches can be used in order to look at differences in preferences. For example, we could have simply run separate econometric models for different sub-groups of respondents. If we also ran a pooled model, then a likelihood ratio (LR) test could then be used to establish whether preferences differed in a statistically significant manner, comparing the sub-groups with the pooled sample. However, a limitation of this approach is that it would not confer information about whether preferences for a given attribute (relating to a given potential transplant allocation criterion) differed in a statistically significant manner. That is why we decided to use interaction dummy variables to establish whether preferences differed for certain attributes comparing healthcare professionals, donors / relatives of deceased donors, and carers with patients. Using interaction dummy variables you can establish whether preferences for a given attribute are statistically significant compared to the patient (base) group. This same approach was used to compare ethnic minority versus non-ethnic minority patients.

In Chapter 1 (section D) the findings of a review of DCEs relating to health are summarized. It is pointed out (section D.7.2) that in order to present DCE results in a form which policy makers can readily use, some kind of summary measure can be deployed. In the past the main ones used related to ‘per WTP unit’ or some kind of monetary welfare measure. However, as discussed in Chapters 3 and 4 there are a range of methodological problems associated with such an approach. In addition, when it comes to renal transplantation, we

considered that the use of a monetary attribute would be inappropriate. This is because, if a transplant has a reasonable probability of being successful, it is usually far less expensive to provide a transplant than for a patient to remain on dialysis (which is very expensive). Therefore, in the context of renal transplantation the scarce resource is the organ. Moreover, in the UK legislation prohibits the sale of organs, so it might be considered unethical to pose questions designed to get people to value kidneys in monetary terms.

In practice, in the UK people who are on the transplant list have to wait to see if a suitable kidney becomes available. Thus they operate in an environment in which there is rationing by queues. Therefore, we decided to summarise our findings in terms of a measure of marginal rate of substitution (MRS) which relates differences in other attributes (potential transplant priority criteria) to waiting time for renal transplants. Once a summary outcome measure of this sort is adopted, it begs the question (if you are trying to establish if preferences vary between key stakeholder groups) of whether any differences in MRS are statistically significant differences in MRS between stakeholder groups for attributes. We have therefore gone further than most DCE studies published to date, in that we have applied Wald tests (implemented using the econometric package STATA) in order to establish whether MRS for attributes differs between stakeholder groups.

2. Background.

In the United Kingdom (UK), in January 2011, 6,610 patients were awaiting renal transplantation (this figure has risen 8% annually since 2004), and in 2009-10, 1,482 patients received a deceased donor transplant, and 1,038 received live donor transplants (NHS (Blood and transplant 2011). This growing imbalance between demand and supply led to the 2008 Organ Donation Taskforce Report (Department of Health 2008) which outlined

initiatives to increase organ supply by 50% within 5 years. Despite this, demand will continue to outstrip supply despite increased supply, so criteria do need to be in place to allocate the limited supply of kidneys available for transplantation. Efficiency requires organs be transplanted to patients who will derive the greatest health benefit. Equity of access concerns, however, may conflict with efficiency ones. Thus, patients waiting a long time may be given a transplant on equity grounds, even if someone else who has not waited as long would obtain greater health benefit from transplantation.

UK transplant policy was previously re-appraised in 2006. The existing policy was recognized to disadvantage those with less common tissue types and blood groups, especially ethnic minorities (Higgins, West et al. 1997). At the same time, populations such as African Caribbeans and South Asians have a 3-4 times greater risk of end stage renal disease (Department of Health 2008), related to the higher prevalence of type 2 diabetes (Raleigh 1997). The increased risk of renal disease is also linked to an increased risk of hypertension (Norris, Tareen et al. 2008) combined with increased cardiovascular disease (Cappuccio, Oakeshott et al. 2002). Moreover, ethnic minorities donate fewer organs (Cappuccio, Oakeshott et al. 2002) so patients are less able to obtain closely matched transplants.

The 2006 re-appraisal reduced priority attached to HLA matching, allowing consideration of other criteria (Koenne 2002). The resulting guidelines (NHS Blood and Transplant 2006) recommended that more priority should be given to long waiters and to paediatric and younger adult recipients. Research in the USA and Australia had indicated that such changes would be acceptable to professionals and patients (Louis, Sankar et al. 1997; Browning and Thomas 2001).

At the same time, DCEs were increasingly being used to address healthcare priority setting issues in primary care (Rubin, Bate et al. 2006) and secondary care (Allepuz, Espallargues et al. 2008; Youngkong, Baltussen et al. 2010). Some DCE transplantation work had been published, including work assessing factors influencing willingness to donate body parts (Bennett and Savani 2004) and a UK DCE to establish priorities for liver transplantation (Ratcliffe and Buxton 1999; Ratcliffe 2000) . In renal transplantation, the first published DCE findings emanated from my own study which was conducted in the UK (Clark, Gumber et al. 2009). This publication looked at diversity issues and assessed whether patient preferences varied by ethnicity and gender (the subject of Chapter 6 of this thesis). In contrast, the present chapter considers the preferences of other stakeholder groups including renal healthcare professionals, renal carers, and live donors / relatives of deceased donors, alongside those of patients. Also, it examines how patient preferences vary between ethnic minority (including white ethnic minorities) and non-ethnic minority patients. In the following chapter (Chapter 6), and in our published paper (Clark, Gumber et al. 2009), we examine differences in preferences for non-white patients vs. all other patients, and South Asian patients vs. all other patients. More recently, there has also been some renal DCE research relating to patient and healthcare professional preferences for chronic kidney disease (CKD) care more generally (including kidney transplantation) in Canada (Davison, Kromm et al. 2010).

Unlike the general public (who lack personal experience of renal disease, and are unlikely to ever require kidney transplants unless they suffer from hypertension, diabetes, cardiovascular disease or obesity), the stakeholder groups considered in this chapter are likely to have a more direct interest in transplant decisions. For example, patients, carers and healthcare professionals all have an interest in who is prioritized for transplantation, either because they have renal disease (patients) or care for those with renal disease (renal carers / renal

healthcare professionals). Moreover, donors or relatives of deceased donors are also clearly concerned to ensure that kidneys are appropriately allocated. Therefore, these four stakeholder groups can be expected to be generally better informed than the general public about transplantation.

3. Methods.

This DCE involves respondents making choices, about which one of two hypothetical transplant recipients (differing in characteristics) should receive a kidney. DCE respondents trade-offs are established so weightings given to different recipient characteristics (attributes) are quantified. The pilot study began in 2005; the main study in 2006. Final data analysis was during 2007–09.

3.1. Pilot exercise.

We developed a pilot DCE questionnaire using SPEED (Bradley 1991). Attributes and level selection, was mainly informed by discussions with clinicians. We paired choices generated by SPEED to minimize attribute overlap and level imbalance (Huber and Zwerina 1996). We interviewed 60 respondents (who completed questionnaires and ranked potential attributes) to inform attribute and level selection. Respondents included 41 patients including 8 ethnic minorities, 16 healthcare professionals, 1 donor, 1 carer, and a renal Consultants secretary). They completed a DCE questionnaire, and ranked attributes (as described in the questionnaire and written on cards) in priority order. Pilot respondents could also suggest other potential attributes, and details of these were written on cards. They then ordered cards in order of priority. Most respondents (n = 56) came from the University Hospital, Coventry, including 4 ethnic minorities. Another 4 ethnic minority patients came from Ealing NHS Trust

to boost minority responses. Pilot DCE attributes and levels included: Waiting time (levels: 1 month, 2 years, or 10 years); tissue match (levels: non-favourable, favourable, and perfect,); employment status (levels: unemployed, part-time, or full time); number of dependent children or adults (levels: 0, 1, or 4); extra years of life expectancy (levels: 1, 5 or 12 years); recipient age (levels: 20, 45, or 70 years); and other recipient diseases (levels: healthy accept for kidney disease, kidney disease plus a condition affecting activities [asthma], and kidney disease plus a condition affecting daily activities [severe arthritis]).

3.2. Attributes and levels - final DCE.

We analyzed the pilot data using Random Effects Probit and all the attributes (with the exception of the employment status attribute) proved significant at the 5% level. Early during piloting respondents expressed disquiet about the employment attribute, arguing it represented unwarranted discrimination against the retired or those unemployed because of illness. We therefore asked respondents whether this should be an attribute, and most respondents said 'no', so we dropped it. However findings from attribute rankings suggested the following warranted inclusion. Most respondents thought people with adult and child dependents ought to be prioritized, so the dependents attribute was amended to include adults. Age was considered relevant but the recipient age ceiling was reduced to 65, because clinicians suggested transplantation was unlikely amongst over 65s. The separate life expectancy and other recipient diseases attributes although highly ranked, resulted in unrealistic DCE scenarios. One pairwise choice resulted in a choice between a 70 year old with severe arthritis with 12 years life expectancy, and a 45 year old without co-morbidities having shorter life expectancy. The comparison did not make sense, since you expect a 45 year old without co-morbidities to have longer life expectancy than a 70 year old with severe arthritis. So we replaced the life expectancy attribute with one relating to whether a potential

recipient had diseases predominantly affecting life expectancy. This resulted in more realistic scenarios, improving DCE identification and efficiency properties.

Other highly ranking attributes included patient compliance / whether illness was self-inflicted. This assumed a higher ranking than it normally would because the pilot exercise arose when George Best (a former international football player) was dying after liver transplantation because of on-going alcohol misuse. We wanted to prevent health professionals' preferences over-riding respondents preferences, but healthcare professionals rightly pointed out kidney disease rarely arises because of alcohol misuse, so the issue was not pertinent. Also healthcare professionals pointed out those who are thought likely to abuse their bodies or be non-compliant would not be transplanted. So we excluded this attribute. Table i indicates final attributes and levels.

DCE attributes and levels were explained in the questionnaire preamble. We expected respondents to prioritize those waiting longer for a transplant on equity grounds, so anticipated a positive coefficient on a year's less waiting time ('wait'). The questionnaire preamble explained transplant survival rates obtained from UK Transplant were contingent upon donor / recipient tissue match, and said categories included perfect matches (90% 12 month transplant survival rate when all 6 tissue types match); favourable matches (89% 12 month transplant survival rate when 4-5 tissue types match); or non-favourable matches (86% 12 month survival rate when less than 4 tissue types match). On efficiency grounds we thought respondents would generally prefer to transplant to recipients with the highest chance of success, so improvements in kidney survival would be positively valued, but, some ethnic minority groups might not have this preference if there are a lack of organs closely matching their own. We considered that recipients with child or adult dependents or more dependents would be prioritized because more people benefit from recipients improved health if they care

for others, so expected a positive coefficient. DCE attributes and levels were explained in the questionnaire preamble. We expected respondents to prioritize those waiting longer for a transplant on equity grounds, so anticipated a positive coefficient on a year's less waiting time ('wait').

Table i: Final attributes and levels.

Attribute	Variable name	Levels	Interpretation of coefficients.
Time spent awaiting transplantation	Wait	1 month, 2 years, and 10 years.	Indirect utility of each 1 year reduction in transplant recipient waiting time.
Tissue type matching	Tiss	Non-favourable match: 86% average kidney survival rate post-transplant. Favourable match: 89% average kidney survival rate post-transplant. Perfect match: 90% average kidney survival rate post-transplant.	Indirect utility of prioritizing people for each 1% improvement in kidney survival.
How many child or adult dependents recipients have	Dep	None, 1, or 4 dependents.	Indirect utility of each additional dependent.
Recipient age	Age	20 years, 45 years, and 65 years	Indirect utility for each 1 year reduction in recipient age.
Diseases predominantly affecting life expectancy	dis1 dis2	No disease affecting life expectancy (other than Kidney disease) vs. moderate disease (uncontrolled hypertension or obesity) & Kidney disease. Moderate disease (uncontrolled hypertension or obesity) affecting life expectancy vs. severe disease (heart attack, stroke, or diabetes with complications).	Indirect utility of having no rather than moderate disease predominantly affecting life expectancy. Indirect utility of having moderate disease rather than severe disease predominantly affecting life expectancy.
Diseases predominantly affecting quality of life	ill1 ill2	No disease affecting quality of life (other than Kidney disease) vs. moderate disease (mild asthma). Moderate disease (mild asthma) affecting quality of life vs. severe disease (severe arthritis).	Indirect utility of having no disease rather than a moderate disease predominantly affecting quality of life. Indirect utility of having a moderate disease rather than a severe disease predominantly affecting quality of life.

The questionnaire preamble explained transplant survival rates obtained from UK Transplant were contingent upon donor / recipient tissue match, and said categories included perfect matches (90% 12 month transplant survival rate when all 6 tissue types match); favourable matches (89% 12 month transplant survival rate when 4-5 tissue types match); or non-favourable matches (86% 12 month survival rate when less than 4 tissue types match). On efficiency grounds we thought respondents would generally prefer to transplant to recipients with the highest chance of success, so improvements in kidney survival would be positively valued, but, some ethnic minority groups might not have this preference if there are a lack of organs closely matching their own. We considered that recipients with child or adult dependents or more dependents would be prioritized because more people benefit from recipients improved health if they care for others, so expected a positive coefficient. All other things being equal you expect older patients to benefit less from a transplant because they have a lower life expectancy, so the coefficient on reductions in recipient age should be positive. We anticipated respondents for efficiency reasons would prioritize those with no diseases predominantly affecting life expectancy over those with moderate diseases predominantly affecting life expectancy, and those with moderate diseases predominantly affecting life expectancy over those with severe diseases predominantly affecting life expectancy, so expected positive coefficients. Likewise we anticipated respondents would value prioritizing those with no disease predominantly affecting quality of life to those with moderate diseases predominantly affecting quality of life, and those with moderate diseases predominantly affecting quality of life to those with severe diseases predominantly affecting quality of life, so expected positive coefficients.

3.3 Development of final DCE.

We wanted to force a choice, so used a binary dependent model, because in reality transplant decisions have to be made, and medical professionals face a forced choice when allocating kidneys because of donor scarcity. Moreover pilot interviews revealed many respondents felt uncomfortable with deciding who to transplant. Therefore having a 'cannot decide' option would have triggered such responses from respondents who in reality were not indifferent, so we forced a choice. We could have allowed choices between more than 2 potential recipients using a multinomial model, or had more attributes and levels, but this would complicate decision making (Amaya-Amaya, Gerard et al. 2008). Moreover, many renal patients suffer from fatigue, so we wanted to avoid complexity, because when complexity increases respondents may be more inclined to use simplifying heuristics (Lloyd 2003) compromising response reliability.

The final DCE design did not use SPEED, but was again an Orthogonal Main Effects Plan (OMEP) design involving independent valuation of attributes. To ensure a perfectly orthogonal design we used an OMEP design supplied by leading DCE designers (Street, Burgess et al. 2005) improving efficiency. We blocked 18 choices into 2 blocks of 9 questions (versions A and B) to reduce respondent fatigue; otherwise the patient questionnaire would have been too long. Respondents chose between transplanting patient A or B. For example for one choice patient A waited 2 years; had an 89% chance of 1 year transplant success; had 4 dependents; was 20 years; had severe diseases predominantly affecting life expectancy (heart attack, stroke, or diabetes with complications); but no diseases predominantly affecting quality of life except Kidney disease. Patient B waited 10 years; had a 90% chance of 1 year transplant success; had no dependents; was 45 years; had no diseases predominantly affecting life expectancy except kidney disease; and had moderate

disease affecting quality of life (mild asthma). See appendix E which contains copies of the questionnaires which were tailored for the 4 main different stakeholder groups

3.4. Questionnaire distribution.

The National Kidney Federation included a flyer and freepost envelope in 'Kidney Life' (circulation c.20,000) inviting patients, carers, donors, or healthcare professionals to request questionnaires. Members of the British Organ Donor Society had questionnaires posted to them. We sent questionnaires to healthcare professionals listed in UK transplants service directory, and targeted transplanting units with transplant coordinators or transplant physicians. To increase ethnic minority responses we provided translated questionnaires. A reputable translation organization was used to translate questionnaires into Punjabi, Hindi, Bengali, Gujarati, and Urdu. The bilingual researcher administering the questionnaires upon non-English speaking patients then checked the questionnaires translations accuracy, and chased ethnic minority patient responses obtaining 18 additional responses from Ealing NHS Trust, and 5 from University Hospital, Coventry.

3.5. Econometric / statistical analysis.

We used Random Effects Probit (model 1), to establish stakeholder preferences.

$$Y_{ij} = \beta_0 + \beta_1 wait_{ij} + \beta_2 tiss_{ij} + \beta_3 dep_{ij} + \beta_4 age_{ij} + \beta_5 dis1_{ij} + \beta_6 dis2_{ij} + \beta_7 ill1_{ij} + \beta_8 ill2_{ij} + \beta_9 D_{cij} + \beta_{10} D_{cij} wait_{ij} + \beta_{11} D_{cij} tiss_{ij} + \beta_{12ij} D_c dep_{ij} + \beta_{13ij} D_c age_{ij} + \beta_{14} D_{cij} dis1_{ij} + \beta_{15} D_{cij} dis2_{ij} + \beta_{16} D_{cij} ill1_{ij} + \beta_{17} D_{cij} ill2_{ij} + \beta_{18} D_{dij} + \beta_{19} D_{dij} wait_{ij} + \beta_{20} D_{dij} tiss_{ij} + \beta_{21} D_{dij} dep_{ij} + \beta_{22} D_{dij} age_{ij} + \beta_{23} D_{dij} dis1_{ij} + \beta_{24} D_{dij} dis2_{ij} + \beta_{25} D_{dij} ill1_{ij} + \beta_{26} D_{ijd} ill2_{ij} + \beta_{27} D_{hij} + \beta_{28} D_{hij} wait_{ij} + \beta_{29} D_{hij} tiss_{ij} + \beta_{30} D_{hij} dep_{ij} + \beta_{31} D_{hij} age_{ij} + \beta_{32} D_{hij} dis1_{ij} + \beta_{33} D_{hij} dis2_{ij} + \beta_{34} D_{hij} ill1_{ij} + \beta_{35} D_{hij} ill2_{ij} + \mu_i + \xi_{ij} \quad \text{(Model 1)}$$

The term Y_{ij} is a binary dependent variable, from individuals i = 1...m, for observations j = 1...n_i. Observations n_i vary because the i individuals do not all complete every pairwise choice (some respondents do not answer all choices), μi is the random effects error term (which allows for multiple responses from i respondents), and ε_{ij} is the probit error term for individuals i for j observations.Variables are defined in table i. Prefixes on dummy variables (D_c, D_d, and D_h), indicate carer, donor, and healthcare professional preferences respectively, D_c = 1 for carers, 0 otherwise; D_d = 1 for donor / relatives of deceased donors, 0 otherwise; and D_h = 1 for healthcare professionals, 0 otherwise. Model 1 establishes whether carers, donors, or healthcare professional preferences differ from patients. If β_9, $\beta_{18,}$ or $\beta_{27,}$ are significant this indicates general non-attribute specific differences in carer, donor, or healthcare professional preferences compared to patients. Other dummy variables are interaction dummies. If any carer coefficients ($\beta_{10....}\beta_{17}$) are significant, this indicates that preferences for associated attribute(s) differ between carers and patients. If dummies for donors ($\beta_{19....}\beta_{26}$) are significant, it indicates preferences between donors and patients for the attribute(s) differ. Likewise, if interaction dummies for healthcare workers ($\beta_{28....}\beta_{35}$) are significant, it indicates different preferences between them and patients for attribute(s).

Model 2 compares ethnic and non-ethnic minority patient preferences. The ethnic minority patient category included all patients in an ethnic category except 'White British.' Y_{ij}, μ_i, and ξ_{ij} are as previously defined, D_E is a dummy variable, D_E = 1, for ethnic minorities, 0 otherwise.

$$Y_{ij} = \beta_0 + \beta_1 wait_{ij} + \beta_2 tiss_{ij} + \beta_3 dep_{ij} + \beta_4 age_{ij} + \beta_5 dis1_{ij} + \beta_6 dis2_{ij} + \beta_7 ill1_{ij} + \beta_8 ill2_{ij} + \beta_9 D_{Eij} + \beta_{10} D_{Eij} wait_{ij} + \beta_{11} D_{Eij} tiss_{ij} + \beta_{12} D_{Eij} dep_{ij} + \beta_{13} D_{Eij} age_{ij} + \beta_{14} D_{Eij} dis1_{ij} + \beta_{15} D_{Eij} dis1_{ij} + \beta_{16} D_{Eij} ill1_{ij} + \beta_{17} D_{Eij} ill2_{ij} + \mu_i + \xi_{ij}$$

(Model 2)

Model 2 establishes whether preferences differ between ethnic and non-ethnic minority patients. If β_9 is significant it suggests non-attribute specific differences in preferences between ethnic and non-ethnic minority patients. If ethnic minority interaction dummies ($\beta_{10....}\beta_{17}$) are significant, it indicates preferences differ between ethnic minorities and non-ethnic minorities for significant associated attribute(s).

3.6. Statistical methods for Marginal Rate of Substitution (MRS).

MRS indicates the ratio of changes in other attributes to changes in waiting times (see table ii) We used the Delta method (Wooldridge 2002) using command 'nlcom' in STATA, for 95% confidence intervals, to establish statistical significance.

Wald tests using 'testnl' in STATA established whether MRS differed significantly between groups, comparing patients with carers, donors, and healthcare professionals (model 1), and ethnic minority vs. non-ethnic minority patients (model 2). Thus to establish (model 2) whether preferences for tissue matching differed between ethnic and non-ethnic minorities, the hypothesis is $\beta_2 / \beta_1 = (\beta_2 + \beta_{11}) / (\beta_1 + \beta_{10})$ i.e. was tissue match MRS for non-ethnic and ethnic minorities identical ($p \leq 0.05$ indicates a difference at the 5% level).

Table ii: Calculating MRS – Valuing attributes compared to a 1 year difference in wait time.

Model 1 Variable	**Patient MRS**	**Model 1 Variable**	**Carer MRS**
Wait		**Wait**	
Tiss	β_2 / β_1	**Tiss**	$(\beta_2 + \beta_{11}) / (\beta_1 + \beta_{10})$
Dep	β_3 / β_1	**Dep**	$(\beta_3 + \beta_{12}) / (\beta_1 + \beta_{10})$
Age	β_4 / β_1	**Age**	$(\beta_4 + \beta_{13}) / (\beta_1 + \beta_{10})$
dis1	β_5 / β_1	**dis1**	$(\beta_5 + \beta_{14}) / (\beta_1 + \beta_{10})$
dis2	β_6 / β_1	**dis2**	$(\beta_6 + \beta_{15}) / (\beta_1 + \beta_{10})$
ill1	β_7 / β_1	**ill1**	$(\beta_7 + \beta_{16}) / (\beta_1 + \beta_{10})$
ill2	β_8 / β_1	**ill2**	$(\beta_8 + \beta_{17}) / (\beta_1 + \beta_{10})$
Model 1 Variable	**Donor MRS**	**Model 1 Variable**	**Healthcare professional MRS**
Wait		**Wait**	
Tiss	$(\beta_2 + \beta_{20}) / (\beta_1 + \beta_{19})$	**Tiss**	$(\beta_2 + \beta_{29}) / (\beta_1 + \beta_{28})$
Dep	$(\beta_3 + \beta_{21}) / (\beta_1 + \beta_{19})$	**Dep**	$(\beta_3 + \beta_{30}) / (\beta_1 + \beta_{28})$
Age	$(\beta_4 + \beta_{22}) / (\beta_1 + \beta_{19})$	**Age**	$(\beta_4 + \beta_{31}) / (\beta_1 + \beta_{28})$
dis1	$(\beta_5 + \beta_{23}) / (\beta_1 + \beta_{19})$	**dis1**	$(\beta_5 + \beta_{32}) / (\beta_1 + \beta_{28})$
dis2	$(\beta_6 + \beta_{24}) / (\beta_1 + \beta_{19})$	**dis2**	$(\beta_6 + \beta_{33}) / (\beta_1 + \beta_{28})$
ill1	$(\beta_7 + \beta_{25}) / (\beta_1 + \beta_{19})$	**ill1**	$(\beta_7 + \beta_{34}) / (\beta_1 + \beta_{28})$
ill2	$(\beta_8 + \beta_{26}) / (\beta_1 + \beta_{19})$	**ill2**	$(\beta_8 + \beta_{35}) / (\beta_1 + \beta_{28})$
Model 2 Variable	**Non-ethnic minority MRS**	**Model 2 Variable**	**Ethnic minority MRS**
Wait		**Wait**	
Tiss	β_2 / β_1	**Tiss**	$(\beta_2 + \beta_{11}) / (\beta_1 + \beta_{10})$
Dep	β_3 / β_1	**Dep**	$(\beta_3 + \beta_{12}) / (\beta_1 + \beta_{10})$
Age	β_4 / β_1	**Age**	$(\beta_4 + \beta_{13}) / (\beta_1 + \beta_{10})$
dis1	β_5 / β_1	**dis1**	$(\beta_5 + \beta_{14}) / (\beta_1 + \beta_{10})$
dis2	β_6 / β_1	**dis2**	$(\beta_6 + \beta_{15}) / (\beta_1 + \beta_{10})$
ill1	β_7 / β_1	**ill1**	$(\beta_7 + \beta_{16}) / (\beta_1 + \beta_{10})$
ill2	β_8 / β_1	**ill2**	$(\beta_8 + \beta_{17}) / (\beta_1 + \beta_{10})$

4. Results.

4.1.Sample characteristics.

Table iii indicates respondent characteristics. UK Renal Registry data (Byrne, Ford et al. 2008; Byrne, Steenkamp et al. 2008) was used to assess patient sample representativeness. Of the 895/ 908 patients indicating ethnicity, 799 / 895 patients (89.3%) were white (British),

and 27 / 895 (3%) were white ethnic minorities, so 92.3% are white. UK incidence data (Byrne, Ford et al. 2008) suggested 79.7% of renal patients are white, so whites are over-represented in our survey. Overall, 69 / 895 (7.7%) patients indicating ethnicity were non-white, compared with a 20.3% incidence rate (Byrne, Ford et al. 2008), 50 / 69 non-white patients were South Asians (5.6% of those indicating ethnicity) compared to a 10.5% incidence (Byrne, Ford et al. 2008). 508 /908 patients (55.9%) were male, 397 / 908 (43.7%) were female, 3 / 908 (0.3%) did not say. Graphically presented Renal Registry data (Byrne, Ford et al. 2008) reassuringly indicated slightly higher proportions of male than female patients across age groups. Average sample patient age was 54.88 years (median 57 years), and Renal Registry data median age (57.3 years) was virtually identical (Byrne, Steenkamp et al. 2008).

Table iii: Sample characteristics.

	Patients (n = 908)	Carers (n=41)	Donors (n =48)	Healthcare workers (n=113)
AGE				
Mean age	54.88 years	52.37 years	54.67 years	43.23 years
GENDER				
Male	508 (55.9%)	10 (24.4%)	14 (29.2%)	51 (45.1%)
Female	397 (43.7%)	31 (75.6%)	34 (70.8%)	61 (54.0%)
Not indicated	3 (0.3%)	0 (0%)	0 (0%)	1 (0.9%)
ETHNICITY				
White (British)	799 (88%)	38 (92.7%)	44 (91.7%)	89 (78.8%)
White ethnic minorities	27 (2.9%)	1 (2.4%)	1 (2.1%)	9 (8%)
Non-white ethnicity (excluding Asians)	19 (2.1%)	1 (2.4%)	0 (0%)	2 (1.8%)
Non-white ethnicity (Asians)	50 (5.5%)	1 (2.4%)	1 (2.1%)	9 (8%)
Not indicated	13 (1.4%)	0 (0%)	2 (4.2%)	4 (3.5%)
DEPENDENT CHILDREN				
0	755 (83.1%)	33 (80.5%)	36 (75%)	51 (45.1%)
1	72 (7.9%)	2 (4.9%)	5 (10.4%)	22 (19.5%)
2	49 (5.4%)	5 (12.2%)	2 (4.2%)	26 (23.0%)
3	12 (1.3%)	0 (0%)	3 (6.3%)	9 (8.0%)
> 3	7 (0.8%)	1 (2.4%)	1 (2.1%)	3 (2.7%)
Not indicated	13 (1.4%)	0 (0%)	1(2.1%)	2 (1.8%)
DEPENDENT ADULTS				
0	750 (82.6%)	16 (39.0%)	39 (81.2%)	98 (86.7%)
1	121 (13.3%)	17 (41.5%	6 (12.5%)	11 (9.7%)
2	17 (1.9%)	6 (14.6%)	2 (4.2%)	3 (2.7%)
> 2	8 (0.9%)	1 (2.4%)	0 (0%)	0 (0%)
Not indicated	12 (1.3%)	1 (2.4%)	1 (2.1%)	1 (0.9%)

The patient sample comprised: 468 / 908 (51.5%) with successful transplants; 118 / 908 (13%) whose transplant failed; 279 / 908 (30.7%) awaiting transplants (average wait 22.6 months). Some patients whose transplant failed also appeared as awaiting transplantation; 237 / 908 (26.3%) had dialysis without transplantation; and 57 / 908 (6.3%) had kidney disease, not requiring dialysis. Renal Registry prevalence data (Byrne, Steenkamp et al.

2008) suggests 46.9% of patients have successful transplants (close to our figure). There is no incidence / prevalence data for other patient categories. Amongst non-whites (including Asians) figures are 18 / 69 patients (26%) with successful transplants; 10 / 69 (14.5%) whose transplant failed; 35 / 69 patients (50.7%) awaiting a transplant on dialysis (average wait: 21.45 months); and 3 / 69 (4.3%) with kidney disease, not requiring dialysis. Unfortunately renal registry data (Byrne, Steenkamp et al. 2008) does not indicate ethnicity. However, lower percentage figures for successes, and higher transplant failures figures are expected (ethnic minorities are less likely to be closely matched).

There were 48 donor responses (21 live donors, and 27 deceased donor relatives).
Healthcare professionals comprised: 9 renal surgeons; 37 renal physicians; 17 transplant co-ordinators; 31 nurses; 9 clinical scientists; 1 GP; 1 dietician; 1 network manager; 1 transplant scientist; 1 medical student; 1 transplant immunologist; 1 tissue typer; 1 clinical audit manager; 1 renal technologist; and a pathologist.

4.2. Data analysis.

Table iv indicates model 1 results, figures under MRS are 95% confidence intervals (CIs).

Table iv: Model 1: Results, and MRS, for patients, carers, donors, and healthcare workers.

Attribute	Coefficient : Patients	Implied MRS for patients		Coefficient : Dummies for the carer group	Implied MRS for carers	Wald test p-values: Carers vs. patients
Wait	.0443**	1		-.0156	1	
Tiss	.0624**	1.41** (1.08 / 1.74)		-.0407	0.76 (-1.42 / 2.93)	p=0.0024
Dep	.0635**	1.43** (1.17 / 1.69)		-.0585*	0.17 (-1.54 / 1.89)	p<0.0001
Age	.0069**	0.16** (0.12 / 0.19)		.0006	0.26 (-0.03 / 0.56)	p=0.0750
dis1	-.0004	-0.01 (-1.03 / 1.01)		.1205	4.18 (-3.25 / 11.62)	p=0.2965
dis2	.6789**	15.32** (13.45 / 17.20)		-.1971	16.79* (2.69 / 30.89)	p<0.0001
ill1	-.1207**	-2.73** (-1.45 / -4.00)		.1130	-0.27 (-9.55 / 9.01)	p=0.1236
ill2	.1850**	4.18** (3.12 / 5.23)		-.0334	5.28 (-2.27 / 12.83)	p=0.0910
Intercepts	.1208**			-.0034		
Attribute	**Coefficient : Dummies for the donor group**	**Implied MRS for donors**	**Wald test p-values: Donors vs. patients**	**Coefficient : Dummies for the healthcare worker group**	**Implied MRS for healthcare workers**	**Wald test p-values: Healthcare workers vs. patients**
Wait	-.0086	1		-.0039	1	
Tiss	-.0667*	-0.12 (-1.62 / 1.38)	p<0.0001	-.0110	1.27* (0.24 / 2.31)	p=0.0027
Dep	-.0468*	0.47 (-0.79 / 1.73)	p<0.0001	-.0003	1.56** (0.72 / 2.41)	p=0.0017
Age	-.0023	0.13 (-0.05 / 0.31)	p=0.0067	.0127**	0.48** (0.31 / 0.66)	p=0.0300
dis1	.1508	4.22 (-1.38 / 9.81)	p=0.1669	.1823**	4.50** (1.09 / 7.91)	p=0.0265
dis2	-.2676*	11.54** (2.90 / 20.17)	p<0.0001	.1056	19.42** (12.71 / 26.14)	p<0.0001
ill1	.0520	-1.93 (-8.66 / 4.80)	p=0.2301	.0501	-1.75 (-5.74 / 2.24)	p=0.1048
ill2	.0245	5.87* (0.18 / 11.57)	p=0.1942	.1790*	9.01** (5.44 / 12.59)	p=0.9818
Intercepts	-.112			.0844		

*: Significant at 5% level; **:Significant at 1% level

Coefficients interpretation is indicated in table i. MRS in table iv indicates indirect utility values for changes in attributes (for direction of change see table i) relative to indirect utility values for prioritizing a recipient waiting an extra year for transplantation (table ii indicates MRS formulae). Measures of 'goodness of fit' for model 1 show that 63.06% of actual values are predicted by the model, and Mc Faddens $R^2 = 0.1088$. A likelihood ratio test to test for the joint significance of the dummy variables has 27 degrees of freedom with a critical value of 40.11, compared with λ=71.90, so the dummy variables are jointly significant. The tissue match coefficient (tiss) in table iv indicates the impact of a 1% difference in 12 month kidney survival. Difference in survival rates between a perfect vs. favourable match is 1%, so the MRS figure of 1.41 (table iv) also appears in table v, for 'Prioritizing perfect not non-favourable tissue matches'. Table v indicates MRS for the 'Prioritizing someone with a favourable not non-favourable match' (1.41 x 3 [a 3% difference in kidney survival rate] = 4.23). It also indicates how much respondents value other changes in attributes to a 1 year wait. Moreover 5 year MRS figures are presented in table v. If waiting time increases 5 fold, MRS for a 5 year wait is 1 / 5^{th} of 1 year MRS. Figures under MRS are 95% CIs, CIs for 5 year MRS are 1 / 5^{th} of 1 year CIs.

4.2.1. Patient preferences.

Patients' MRS figures (table v) suggest all other things being equal (ceteris paribus) patients would prioritize recipients with perfect over favourable tissue matches (tiss) more than those waiting an extra year (1 year MRS =1.41, exceeding indirect utility from avoiding a 1 year wait of 1.00). However, if a favourably matching patient waited 5 years longer, they would be a higher priority than the perfect match (MRS = 0.28) <1. Prioritizing someone with a favourable not non-favourable match (ceteris paribus) is valued more than prioritizing someone waiting

for 1 year (MRS = 4.23), but if the other potential recipient waited 5 years longer, prioritizing the longest waiter is optimal (MRS = 0.85) < 1.

Patient respondents prioritize someone with an extra dependent more than waiting a year longer (MRS for 'dep'=1.43), ceteris paribus. However if the other potential recipient waited 5 years longer MRS = 0.29, so prioritizing an extra dependent is a lower priority. Ceteris paribus prioritizing someone 1 year younger is valued less than a 1 year or 5 year reduction in waiting time ('age' 1 year MRS = 0.16, 5 year = 0.03). Patients would not prioritize those with no vs. moderate diseases predominantly affecting life expectancy significantly ('dis 1' is insignificant), but prioritized (dis2) those with moderate rather than severe diseases predominantly affecting life expectancy highly (1 year MRS = 15.32; 5 year MRS = 3.06). So prioritizing someone with a moderate not severe disease predominantly affecting life expectancy is prioritized (MRS >1).

Paradoxically patients prioritized those with moderate not no diseases predominantly affecting quality of life ('ill1' has a 1 year waiting time MRS of -2.73) ceteris paribus, perhaps because many patients have moderate diseases. However, 5 year MRS = -0.55, so long waiters are a higher priority than those with moderate rather than no disease predominantly affecting quality of life. Finally, patients prioritized those with moderate rather than severe diseases predominantly affecting quality of life ('ill2' 1 year MRS = 4.18; 5 year MRS = 0.84) so ceteris paribus, someone with moderate not severe disease predominantly affecting life expectancy would be a higher priority than someone waiting 1 year longer (MRS>1), but lower priority than someone waiting 5 years longer (MRS < 1).

4.2.2. Carer preferences.

Carer results are compromised by sample size (n=41), so MRS was only significant for 1 variable - prioritizing those with dependents (dep). The fact other MRS figures are insignificant may partly be attributable to the sample size. Table iv indicates Wald test results, p-values ≤ 0.05 indicate statistically significant differences in MRS between other stakeholder groups and patients (5% level). They suggest MRS for prioritizing perfect over favourable tissue matches is lower amongst carers than patients (1 year MRS = insignificant vs. 1.41; 5 year MRS = insignificant vs. 0.28); and lower for prioritizing favourable over non-favourable matches (1 year MRS = insignificant vs 4.23; 5 year MRS = insignificant vs. 0.85). Moreover, Wald tests show carer preference for prioritizing those with dependents is less than patients (1 year MRS = insignificant vs. 1.43; 5 year MRS = insignificant vs. 0.29). Wald tests also show carers prioritize those with moderate not severe diseases predominantly affecting life expectancy (dis2) more than patients (1 year MRS = 16.79 vs. 15.32; 5 year MRS = 3.36 vs. 3.06).

4.2.3 Donor family / live donor preferences.

Donor family / live donor findings are also compromised by small sample size (n=48), which may explain why MRS is only significant for 2 variables ('dis2' and 'ill2'). Wald tests suggested donors value tissue match (tiss) less than patients (1 year MRS = insignificant vs. 1.41; 5 year MRS = insignificant vs. 0.28) for perfect not favourable matches, and also value favourable not non-favourable matches less (1 year MRS = insignificant vs. 4.23; 5 year MRS = insignificant vs. 0.85). Moreover, Wald tests suggest donors value prioritizing dependents (dep) less than patients (1 year MRS = insignificant vs. 1.43; 5 year MRS = insignificant vs. 0.29). They also suggest donors value prioritizing the young (age) less than patients (1 year

MRS = insignificant vs. 0.16; 5 year MRS = insignificant vs. 0.03), and indicate donors prioritized those with moderate rather than severe diseases predominantly affecting life expectancy less than patients (1 year MRS = 11.54 vs. 15.32 ; 5 years MRS = 2.31 vs. 3.06). Wald tests do not indicate other differences.

4.2.4. Healthcare professional preferences.

Healthcare professional MRS is significant for 6 / 7 variables and Wald tests suggest healthcare professional preferences vary from patients for 5 / 7 variables. They valued prioritizing those with better tissue matches 'tiss' less than patients (1 year MRS = 1.27 vs. 1.41; 5 year MRS = 0.25 vs. 0.28) for perfect not favourable matches, and also prioritized favourable vs. non-favourable matches less (1 year MRS = 3.82 vs. 4.23 ; 5 year MRS = 0.76 vs. 0.85). Wald tests also suggested healthcare professionals prioritized those with dependents (dep) more (1 year MRS = 1.56 vs. 1.43 ; 5 year MRS = 0.31 vs. 0.29) and younger patients (age) more (1 year MRS = 0.48 vs. 016; 5 year MRS = 0.10 vs. 0.03). They also valued prioritizing (dis1) those with no vs. moderate diseases predominantly affecting life expectancy when patients do not (1 year MRS = 4.50 vs insignificant; 5 year MRS = 0.90 vs. insignificant). Healthcare professionals prioritized (dis2) those with moderate rather than severe diseases predominantly affecting life expectancy more than patients (1 year MRS = 19.42 vs. 15.32; 5 year MRS = 3.88 vs 3.06). There is no evidence healthcare professionals prioritized those with diseases predominantly affecting quality of life differently from patients (Wald tests for ill1 and ill2 are insignificant). The fact that healthcare professionals have statistically significant differences to patients for 5 / 7 variables, suggests that if healthcare professionals' preferences prevail in transplant decision making (likely if they make decisions) it could result in transplantation allocation decisions which inadequately reflect patient preferences.

Table v: Model 1: MRS, for patients, carers, donors and healthcare workers.

Variable	Patient trade-off between variable & 1 year wait	Patient trade-off between variable & 5 years wait	Carers trade-off between variable & 1 year wait	Carers trade-off between variable & 5 years wait
Prioritizing perfect not favourable tissue matches.	1.41** (1.08 / 1.74)	0.28** (0.22 / 0.35)	0.76 (-1.42 / 2.93)	0.15 (-0.28 / 0.59)
Prioritizing favourable not non- favourable tissue matches.	4.23** (3.23 / 5.22)	0.85** (0.65 / 1.05)	2.27 (-4.25 / 8.80)	0.45 (-0.85 / 1.76)
Prioritizing a recipient with dependents – per extra dependent	1.43** (1.17 / 1.69)	0.29** (0.23 / 0.34)	0.17 (-1.54 / 1.89)	0.03 (-0.31 / 0.38)
Prioritizing a younger recipient – per year younger	0.16** (0.12 / 0.19)	0.03** (0.02 / 0.04)	0.26 (-0.03 / 0.56)	0.05 (-.01 / 0.11)
Prioritizing those with 'no' not 'moderate' diseases affecting life expectancy	-0.01 (-1.03 / 1.01)	0.00 (-0.21 / 0.20)	4.18 (-3.25 / 11.62)	0.84 (-0.65 / 2.32)
Prioritizing those with moderate not severe diseases affecting life expectancy	15.32** (13.45 / 17.20)	3.06** (2.69 / 3.44)	16.79* (2.69 / 30.89)	3.36* 0.54 / 6.18)
Prioritizing those with no not moderate diseases affecting QoL	-2.73** (-1.45 / -4.00)	-0.55** (-0.29 / -0.80)	-0.27 (-9.55 / 9.01)	-0.05 (-1.91 / 1.80)
Prioritizing those with moderate not severe diseases affecting QoL	4.18** (3.12 / 5.23)	0.84** (0.62 / 1.05)	5.28 (-2.27 / 12.83)	1.06 (-0.45 / 2.57)
Variable	**Donors trade-off between variable & 1 year wait**	**Donors trade-off between variable & 5 year wait**	**Healthcare workers trade-off between variable & 1 year wait**	**Healthcare workers trade-off between variable & 5 year wait**
Prioritizing perfect not favourable tissue matches.	-0.12 (-1.62 / 1.38)	-0.02 (0.32 / 0.28)	1.27* (0.24 / 2.31)	0.25 (0.05 / 0.46)
Prioritizing favourable not non- favourable tissue matches.	-0.36 (-4.86 / 4.14)	-0.07 (-0.97/ 0.83)	3.82* (0.72 / 6.93)	0.76* (0.14 / 1.39)

Table v: Model 1 (contd): MRS, for patients, carers, donors and healthcare workers.

Prioritizing a recipient with dependents – per extra dependent	0.47 (-0.79 / 1.73)	0.09 (-0.16 / 0.35)	1.56** (0.72 / 2.41)	0.31** (0.14 / 0.48)
Prioritizing a younger recipient – per year younger	0.13 (-0.05 / 0.31)	0.03 (-0.01 / 0.06)	0.48** (0.31 / 0.66)	0.10** (0.06 / 0.13)
Prioritizing those with no not moderate diseases affecting life expectancy	4.22 (-1.38 / 9.81)	0.84 (-0.28 / 1.96)	4.50** (1.09 / 7.91)	0.90** (0.22 / 1.58)
Prioritizing those with moderate not severe diseases affecting life expectancy	11.54** (2.90 / 20.17)	2.31** (0.58 / 4.03)	19.42** (12.71 / 26.14)	3.88** (2.54 / 5.23)
Prioritizing those with no not moderate diseases affecting QoL	-1.93 (-8.66 / 4.80)	-0.39 (-1.73 / 0.96)	-1.75 (-5.74 / 2.24)	-0.35 (-1.15 / 0.45)
Prioritizing those with moderate not severe diseases affecting QoL	5.87* (0.18 / 11.57)	1.17* (0.04 / 2.31)	9.01** (5.44 / 12.59)	1.80** (1.09 / 2.52)

*: Significant at 5% level; **: Significant at 1% level.

Table vi: Model 2: Ethnic minorities vs. others (96 out of 908 are ethnic minorities).

Variable	Coefficient for non-ethnic minorities	MRS for non-ethnic minorities.	Coefficients on ethnic minority dummy variables	MRS for ethnic minority patients	Wald test p-values
Wait	.0451**	1	-.0061	1	
Tiss	.0698**	1.54** (1.19 / 1.90)	-.0630**	0.17 (-0.82 / 1.17)	p<0.0001
Dep	.0595**	1.32** (1.05 / 1.59)	.0351*	2.42** (1.40 / 3.44)	p=0.2755
Age	.0071**	0.16** (0.12 / 0.20)	-.0011	0.15* (0.03 / 0.27)	p=0.0024
dis1	.0039	0.09 (-0.98 / 1.15)	-.0398	-0.92 (-4.41 / 2.57)	p=0.6014
dis2	.7158**	15.86** (13.87/17.85)	-.3153**	10.25** (4.96 / 15.53)	p<0.0001
ill1	-.1085**	-2.40** (-1.06/-3.74)	-.0903	-5.08* (-0.83/ -9.33)	p=0.9050
ill2	.1773**	3.93** (2.82 / 5.03)	.0647	6.19** (2.51 / 9.88)	p=0.2558
Intercepts	.1269**		-.0510		
Variable	**Non-ethnic minorities trade-off between variable & 1 year wait**	**Non-ethnic minorities trade-off between variable & 5 year wait**	**Ethnic minority trade-off between variable & 1 year wait**	**Ethnic minority trade-off between variable & 5 year wait**	
Prioritizing perfect not favourable tissue matches.	1.54** (1.19 /1.90)	0.31** (0.24 / 0.38)	0.17 (-0.82/1.17)	0.35 (-0.16/0.23)	
Prioritizing favourable not non- favourable tissue matches.	4.64** (3.57 /5.70)	0.93** (0.71 / 1.14)	0.52 (-2.46 / 3.50)	0.10 (-0.49 /0.70)	
Prioritizing a recipient with dependents – per extra dependent	1.32** (1.05/1.59)	0.26** (0.21 / 0.32)	2.42** (1.40/ 3.44)	0.48** (0.28 / 0.69)	
prioritizing a younger recipient – per year younger	0.16** (0.12/ 0.20)	0.03** (0.02 / 0.04)	0.15* (0.03/ 0.27)	0.03* (0.01 / 0.05)	
Prioritizing those with no not moderate diseases affecting life expectancy	0.09 (-0.98/1.15)	0.02 (-0.20 / 0.23)	-0.92 (-4.41/2.57)	-0.18 (-0.88 / 0.51)	
Prioritizing those with moderate not severe diseases affecting life expectancy	15.86** (13.87/17.85)	3.17** (2.77 / 3.57)	10.25** (4.96/15.53)	2.05** (0.99 / 3.11)	
Prioritizing those with no not moderate diseases affecting QoL	-2.40** (-1.06 / -3.74)	-0.48** (-0.21/ -0.75)	-5.08* (-0.83 / -9.33)	**-**1.02* (-0.17 / -1.87)	
Prioritizing those with moderate not severe diseases affecting QoL	3.93** (2.82 / 5.03)	0.79** (0.56 / 1.01)	6.19** (2.51 / 9.88)	1.24** (0.50 / 1.98)	

*: Significant at the 5% level; **: Significant at the 1% level..

4.2.5. Ethnic minority patient preferences.

Measures of 'goodness of fit' for model 2 indicate 62.09% actual values are predicted by the model, and Mc Faddens $R^2 = 0.133$. A likelihood ratio test for the significance of the dummy variables has $\lambda = 35.83$, which compares with a critical value for 9 degrees of freedom of 16.92, so the dummy variables are jointly significant. Table vi compares ethnic and non-ethnic minority patients. Coefficients are defined in table i, and MRS specified in table ii. Overall 3 dummy variables (tiss, dep and dis2) are significant, but Wald tests suggest more variation. Wald tests suggest the following differences. Ethnic minorities do not prioritize recipients with better tissue matches, (tiss) but non-ethnic minorities do (1 year MRS = insignificant vs. 1.54; 5 years MRS = insignificant vs. 0.31 for perfect rather than non-favourable matches). For favourable rather than non-favourable matches only non-ethnic minorities valued favourable matches significantly (1 year MRS = insignificant vs. 4.64; 5 years MRS = insignificant vs. 0.93), perhaps because ethnic minorities are disadvantaged if a close tissue match is required, due to a lack of ethnic minority donors.

Wald test results indicate MRS for prioritizing younger (age) rather than older recipients differs marginally between ethnic minority and non-ethnic minority patients (1 year MRS = 0.15 vs. 0.16; 5 year MRS = 0.03 vs. 0.03); Wald tests suggest ethnic minority patients valued prioritizing recipients with moderate vs. severe diseases (dis2) predominantly affecting life expectancy less than other patients (1 year MRS = 10.25 vs. 15.86; 5 year MRS = 2.05 vs 3.17), perhaps due to higher prevalence of severe diseases predominantly affecting life expectancy amongst ethnic minorities. Wald tests do not indicate valuation of other attributes varies by ethnicity.

5. Discussion.

This study is unique because although DCEs have been used in relation to liver transplantation on the public (Ratcliffe 2000) and patients (Ratcliffe and Buxton 1999), this is the first application of DCEs exclusively relating to prioritizing renal transplants. Moreover, the extent of comparison between stakeholder respondent groups is unprecedented.

5.1. Summary of patient preferences (and how they differ by ethnicity).

Patients valued prioritizing patients with closer tissue matches, but also valued other factors significantly including prioritizing long waiters; those with child or adult dependents; and younger recipients. Those with moderate rather than severe diseases predominantly affecting life expectancy were a priority, but not those with moderate rather than no diseases predominantly affecting life expectancy. Patients also prioritized those with moderate rather than no diseases predominantly affecting quality of life, and those with moderate rather than severe diseases predominantly affecting quality of life. Ethnic minority patients, unlike non-ethnic minority patients did not value tissue match significantly, and valued prioritizing those with severe rather than moderate disease predominantly affecting life expectancy less than non-ethnic minority patients.

5.2. Summary of Carer preferences.

Some Carer preferences differed from patients. Carers did not value prioritizing those with better tissue matches or dependents significantly, but valued prioritizing those with moderate not severe diseases predominantly affecting life expectancy more than patients. Whilst it is

interesting that carer preferences differ from patients, patient preferences may be considered more important if the objective of transplant policy is to maximise benefits to patients.

5.3. Summary of Donor family / Live donor preferences.

Donor families / live donors also did not value prioritizing better tissue matches significantly, and valued transplants to those with dependents, younger recipients, and those with moderate rather than severe disease predominantly affecting life expectancy, less than patients. Donor family / live donor preferences are important, because without donors transplantation is impossible. In hindsight if we repeated the research again to obtain a larger sample we would target people on the organ donor register as well as donor families / live donors.

5.4. Summary of Healthcare professionals' preferences.

Healthcare professionals' preferences differed from patients; they valued prioritizing better tissue matches less than patients but valued prioritizing those with dependents more. They would prioritize those with no rather than moderate diseases predominantly affecting life expectancy when patients would do the opposite, and severe rather than moderate disease predominantly affecting quality of life more than patients.

5.5 The importance of examining the preferences of different stakeholder groups.

Usually when DCEs are used to address healthcare issues they look at patient preferences. In contrast our study compares preferences across an unprecedented range of different stakeholder groups. The approach allows for comparison of preferences between groups,

and assessment of whether differences are statistically significant. Importantly, findings indicate that if you target DCE questionnaires at different stakeholder groups, preferences may differ. So analyses only eliciting preferences for one group may fail to take into account preference heterogeneity. In transplantation establishing that preferences vary between groups is important. Healthcare professionals usually ultimately make transplantation decisions. If they based decisions upon their own preferences they would prioritize those with better tissue matches less than patients, and those with dependents more than patients, and those with moderate rather than severe diseases predominantly affecting quality of life more than patients. They would also prioritize those with no rather than moderate diseases predominantly affecting life expectancy when patients would not. Note the latter choice may be justified if patient preferences are biased because many patient respondents have moderate diseases affecting life expectancy. However, it is less clear that there is a case for healthcare professionals' preferences overriding patients' preferences such that there is less emphasis upon closeness of donor – recipient tissue match; recipients with dependents are prioritized more; and those with moderate not severe diseases predominantly affecting life expectancy are prioritized more. Crucially therefore this analysis suggests that if healthcare professionals base transplant allocation decisions on their preferences this may be in conflict with patient preferences.

5.6. The implications of these findings for the 2006 revisions to UK kidney transplant policy

Our findings are broadly supportive of the 2006 revisions to UK kidney transplant policy, which prioritized long waiters and young adults. Although the analysis presented here shows this is justified, it also suggests other criteria (i.e. prioritizing those with dependents) ought to be considered. Our findings can be compared with other international studies. Australian

based renal research, unlike ours, has adopted a general public perspective (Browning and Thomas 2001). In this case respondents prioritized long waiters and the young, but had a split verdict over whether to prioritize those with children. In the USA, renal research into black Americans (Louis, Sankar et al. 1997) has suggested that allocation based upon HLA matching is considered unfair, although black Americans did not want to receive organs with a lower survival rate (note since this paper's publication graft survival for poorer matches has improved). In Scotland, a Glasgow based patient renal study (Geddes, Rodger et al. 2005) which used a non-DCE scenario approach reported some findings which conflicted with ours (i.e. tissue matching was not a major allocation criterion). However, like our research, preference was given to prioritizing long-waiters (albeit defined by time on dialysis, not on waiting lists). In contrast, more recently Canadian DCE research (Davison, Kromm et al. 2010) has found that respondents preferred to prioritize kidney transplants on the basis of a 'best' match rather than 'first come, first served.' However, their DCE also included attributes relating to organ procurement, and the organization of care for patients with chronic kidney disease in general. So, in contrast to our research, it attempted to value a range of attributes related to CKD and, as such, provided only a very limited analysis of preferences for kidney transplant allocation i.e. there was just one attribute relating to kidney transplant allocation ("How should deceased donor kidneys for transplantation be allocated for transplantation") with just 2 possible levels 'best' match, or 'first come, first served.' Moreover, unlike our DCE study which furnished respondents with data from UK transplants highlighting the fact that the 12 month likelihood of a kidney transplant being successful varied only slightly between non-favourable matches (86% average kidney survival, 12 months post transplantation) and perfect matches (90% average kidney survival 12 months post transplantation), it is not clear from the Canadian paper that similar information was provided to their DCE respondents to ensure an informed response. The findings reported in this chapter, and those reported in our earlier analysis (Clark, Gumber et al. 2009) suggest that both time spent waiting and the

quality of tissue match between donor and recipient are of importance to healthcare workers and to non-ethnic minority patients, but that amongst ethnic minority patients closeness of tissue match is not a significant determinant of patient preferences.

6. Conclusions.

The main conclusions that can be drawn from a methodological point of view are that the use of econometric models which use interaction dummy variables to establish how preferences differ between patients and other key stakeholder groups (healthcare professionals; donor families / live donors; and carers), and to establish how preferences vary between ethnic minority patients and non-ethnic minority patients, seem to work well.

I also conclude that the choice of a summary outcome measure (MRS) which expresses the value of other attributes in terms of transplant waiting times provides an appropriate way of summarizing results (in a manner which is relatively easy to comprehend). Moreover, the use of Wald tests to establish whether MRS for attributes differs significantly between stakeholder groups proved to be useful. The main changes I would make if I were to repeat this study again would be to implement measures to increase the sample of carers and donors available for data analysis. Had I been aware that we would only be able to obtain 41 carer responses through our appeals via Kidney Life, I would have asked patient respondents who requested a questionnaire to supply the name and address of their carer (if they had one). In this way we could have obtained more carer responses for analysis. Similarly, in order to obtain more information about donor preferences, it might have been worth seeking permission to contact people who are registered as kidney donors, in order to ask them to complete a DCE questionnaire.

The main conclusions for policy making purposes are that the findings of our DCE analysis are generally supportive of the changes introduced when the kidney transplant policy was last re-appraised in 2006 (see section 5.6 of this chapter). Although our respondents did not think employment status should be a factor in allocation (which led to its exclusion as a DCE attribute after the pilot stage), having dependents was valued. So the issue of prioritizing those with dependents might be considered when transplant policy is next re-appraised.

Our findings do however raise significant questions around allocation to those from ethnic minority groups, which is why these issues are probed in more detail in the following chapter (Chapter 6). Moreover, the study highlights statistically significant differences in preferences between healthcare professionals and patients. The information on how healthcare professional and patient preferences differ (see sections 5.4 and 5.5 of this chapter) might be of particular use to professionals involved in transplant allocation decisions; it might give them a feel for how the preferences of an average patient differ from those of an average healthcare professional. These findings add to the growing international literature relating to transplant allocation policy, and they ought to be considered when UK renal transplant policy is next re-appraised.

7. References.

Allepuz, A., M. Espallargues, et al. (2008). "Prioritisation of patients on waiting lists for hip and knee arthroplasties and cataract surgery: Instruments validation." BMC Health Serv Res **8**: 76.

Amaya-Amaya, M., K. Gerard, et al. (2008). "Discrete choice experiments in a nutshell." In Using Discrete Choice Experiments to Value Health Care, Edited by Ryan, M., Gerard, K. and Amaya-Amaya, M.(Springer): 13-46.

Bennett, R. and S. Savani (2004). "Factors influencing the willingness to donate body parts for transplantation." J Health Soc Policy **18**(3): 61-85.

Bradley, M. (1991). "Users manual for SPEED version 2.1 stated preference editor and designer." The Hague: Hague Consulting Group.

Browning, C. J. and S. A. Thomas (2001). "Community values and preferences in transplantation organ allocation decisions." Soc Sci Med **52**(6): 853-861.

Byrne, C., D. Ford, et al. (2008). "ESRD incident rates in 2008: national and centre-specific analyses." **Chapter 3, UK Renal Registry Report [http://www.renalreg.org]**.

Byrne, C., R. Steenkamp, et al. (2008). "ESRD prevalent rates in 2008 national and centre-specific analyses." **Chapter 4, UK Renal Registry report [http://www.renalreg.org]**.

Cappuccio, F. P., P. Oakeshott, et al. (2002). "Application of Framingham risk estimates to ethnic minorities in United Kingdom and implications for primary prevention of heart disease in general practice: cross sectional population based study." BMJ **325**(7375): 1271.

Clark, M. D., A. K. Gumber, et al. (2009). "Prioritising patients for renal transplantation? Analysis of patient preferences for kidney allocation according to ethnicity and gender." Diversity in Health and Care, 2009 **6**: 181-191.

Davison, S. N., S. K. Kromm, et al. (2010). "Patient and health professional preferences for organ allocation and procurement, end-of-life care and organization of care for patients with chronic kidney disease using a discrete choice experiment." Nephrol Dial Transplant **25**(7): 2334-2341.

Department of Health (2008). "Organs for transplants: a report from the Organ Donation Taskforce." [http--www.dh.gov.uk-prod_consum_dh-groups-dh_digitalassets-@dh-@en-documents-digitalasset-dh_082120.pdf](16/01/2008).

Geddes, C. C., R. S. Rodger, et al. (2005). "Allocation of deceased donor kidneys for transplantation: opinions of patients with CKD." American Journal of Kidney Diseases **46**(5): 949-956.

Higgins, R. M., N. West, et al. (1997). "Effect of a strict HLA matching policy on distribution of cadaveric kidney transplants to Indo-Asian and white European recipients: regional study." BMJ **315**(7119): 1354-1355.

Huber, J. and K. Zwerina (1996). "The importance of utility balance in efficient choice designs." Journal of Marketing Research **33**(3): 307-317.

Koenne, R. A. (2002). "Should the allocation of cadaveric kidneys for transplantation be based on HLA matching?" Nephrol Dial Transplant **17(5)**: 884-886.

Lloyd, A. J. (2003). "Threats to the estimation of benefit: are preference elicitation methods accurate?" Health Econ **12**(5): 393-402.

Louis, O. N., P. Sankar, et al. (1997). "Kidney transplant candidates' views of the transplant allocation system." J Gen Intern Med **12**(8): 478-484.

NHS Blood and Transplant (2006). "2006 Kidney allocation scheme. ." [http:// www.uktransplant.org.uk-ukt-about_transplants-organ_allocation-kidney_(renal)-renal_organ_sharing_principles-kidney_allocation_scheme_2006-v8.doc.url]

NHS Blood and ransplant (2011). " Weekly Statistics." [http://www.uktransplant.org.uk/ukt/statistics/latest_statistics/latest_statistics.jsp](03/02/2011).

Norris, K. C., N. Tareen, et al. (2008). "Implications of ethnicity for the treatment of hypertensive kidney disease, with an emphasis on African Americans." Nat Clin Pract Nephrol **4**(10): 538-549.

Raleigh, V. S. (1997). "Diabetes and hypertension in Britain's ethnic minorities: implications for the future of renal services." BMJ **314**(7075): 209-213.

Ratcliffe, J. (2000). "Public preferences for the allocation of donor liver grafts for transplantation." Health Econ **9**(2): 137-148.

Ratcliffe, J. and M. Buxton (1999). "Patients' preferences regarding the process and outcomes of life-saving technology. An application of conjoint analysis to liver transplantation." Int J Technol Assess Health Care **15**(2): 340-351.

Rubin, G., A. Bate, et al. (2006). "Preferences for access to the GP: a discrete choice experiment." Br J Gen Pract **56**(531): 743-748.

Street, D. J., L. Burgess, et al. (2005). "Quick and easy choice sets: Constructing optimal and nearly optimal stated choice experiments." International Journal of Research in Marketing **22**(4): 459-470.

Wooldridge, J. M. (2002). "Economic analysis of cross section and panel data." MIT press.

Youngkong, S., R. Baltussen, et al. (2010). "Criteria for priority setting of HIV/AIDS interventions in Thailand: a discrete choice experiment." BMC Health Serv Res **10**: 197.

Chapter 6: Prioritizing patients for renal transplantation? Catering for diversity by analyzing patient preferences for kidney allocation according to ethnicity and gender.

1. Introduction.

In chapter 1 of this thesis (section E1.3), I pointed out that a range of techniques have been used to explore whether respondents' preferences are heterogeneous using DCEs. In this chapter (rather like chapter 5), I apply the interaction dummy variable approach in order to establish whether preferences vary by ethnicity and by gender. It should be noted that the 'non-white ethnic minority' and 'South Asian ethnic minority' categories included in this chapter vary from the ethnic minority patient category of respondents considered in chapter 5 of this thesis (i.e. the white ethnic minority is not included in this chapter).

Of course, it would have been possible to simply run separate regressions models to look at preferences for the different groups. Indeed, many of the DCE studies published during 2009 – 2011 which I identified in the systematic review used this approach (Guimaraes, Marra et al. 2009; Scalone, Mantovani et al. 2009; Tinelli, Ryan et al. 2009; van Til, Stiggelbout et al. 2009; Essers, Dirksen et al. 2010; Torbica and Fattore 2010; van Dam, Hol et al. 2010; Faggioli, Scalone et al. 2011; Pereira, Mulligan et al. 2011; Thrumurthy, Morris et al. 2011). However, a limitation of this approach is that, although you can test for differences in the joint significance of preferences using a likelihood ratio test (comparing a pooled model with two separate models), you cannot establish whether preferences for individual attributes vary in a statistically significant manner.

However, if a model with interaction dummy variables is used, you can also test to establish whether sub-groups of respondents have a statistically significant difference in preferences with respect to certain attributes. Moreover, in this analysis (like the analyses

in chapters 3, 4, and 5) I also take the interaction dummy variable analysis one step further, by testing for differences in the relevant summary ratios between sub-groups. In chapters 3 and 4, we looked for differences in Marginal Willingness to Pay for attributes between sub-groups using Wald tests. In the analyses in chapters 5 and 6, we look at differences in Marginal Rates of Substitution (expressed in terms of transplant waiting times) between sub-groups of respondents using Wald tests.

Most of the material in this chapter has now been published (Clark, Gumber et al. 2009), and should therefore figure in the body of evidence that is used when UK renal transplant policy is next re-appraised (although this chapter updates the literature review presented in our published paper). Having established in chapter 5 that preferences between 'ethnic minority' patients and other patients differed, it was apparent that the equality and diversity issues this raises ought to be probed in more detail. Given that the DCE analysis presented here relates to renal transplantation, it is especially important to establish whether preferences in relation to who should be prioritized for renal transplantation might differ across ethnic minority groups for the following reasons:

- There is a lack of information on the preferences of renal patients generally for different priority criteria for renal transplantation.

- Ethnic minorities, including South Asians, black Africans and African Caribbeans, are more susceptible to renal disease, and there are lower levels of organ donation in these communities.

- Ethnic-minority groups are therefore disadvantaged if allocation is primarily directed towards recipients who can be closely tissue matched with donor organs.

What this chapter adds:

- The analysis provides information about patient preferences and patients' willingness to decide between different priority criteria for transplantation using discrete choice experiments.

- The analysis demonstrates that South Asian and non-white ethnic-minority patients have preferences that differ from those of other patients, in particular that they would not prioritise patients with closer tissue matches or younger respondents. This is in contrast to other patients who are not in these ethnic-minority groups.

- Although there is evidence that preferences for prioritising transplants may differ between male and female patients, gender-related differences in preferences are not particularly pronounced.

2. Background.

Revisions to UK transplant allocation policy in 2006 marked a policy shift towards giving higher priority to people who had been waiting for a long time for transplants, and to young adults, at the expense of emphasising tissue match between donor and recipient. This benefited members of ethnic minorities because of a shortage of donors from some ethnic groups. However, the change was informed by dated research which was not specific to the UK, and which failed to address ethnic or gender-related differences in preferences.

In the UK, in January 2011, 6,610 patients awaited renal transplantation (rising 8% annually since 2004); and in 2009-10, 1,482 received deceased donor transplants, and

1,038 received live donor transplants. The growing imbalance between demand for and supply of transplants had led to the 2008 Transplant Workforce Report (Department of Health 2008), which outlined initiatives to facilitate a 50% increase in cadaveric transplants within five years. Despite this, demand will exceed supply, especially among members of ethnic minority groups. This is because they are more susceptible to certain diseases linked to renal failure necessitating transplants (Raleigh 1997; Churak 2005; Davis and Randhawa 2006; UK Transplant 2006). They are also less likely to obtain closely matched transplants (Higgins, West et al. 1997; UK Transplant 2006). The increased risk among members of ethnic-minority populations, compared with white patients, of developing end-stage renal disease (Churak 2005) is partly related to the higher prevalence of type 2 diabetes. A UK study indicates a prevalence among black African and Caribbean patients that is 3.5 times higher than that among white patients (Raleigh 1997). Type 2 diabetes prevalence is reported to be three to four times higher in UK South Asian adults and the disease may occur a decade earlier than in the white European majority population (Bellary, O'Hare et al. 2008). South Asians are also more susceptible to diabetes and heart failure leading to renal disease (Bennett and Savani 2004). Greater demand for renal transplants in these communities is matched by reduced rates of organ donation (Bennett and Savani 2004). Therefore, systems that prioritize on the basis of donor and recipient tissue matching will disadvantage some ethnic groups.

Improved anti-rejection drugs have reduced the importance of tissue matching in determining transplant success, so other criteria now merit greater consideration (Koene 2002). Moreover, by applying discrete choice experiments (DCEs), it is possible to quantify trade-offs between different priority criteria. A DCE study of this kind has already been undertaken in relation to liver transplantation (Ratcliffe and Buxton 1999; Ratcliffe 2000). In this chapter we report on differences in preferences between non-white ethnic-minority patients and other patients, and between South Asian patients and other patients. We also consider whether preferences vary according to gender.

3. Materials and methods.

This DCE involved respondents making a series of choices about which one of two hypothetical transplant recipients who differ in their characteristics should receive a kidney. Using DCEs, the weight that respondents give to differences in characteristics can be quantified. The steps involved in undertaking this DCE are summarized below.

3.1. Pilot exercise.

We interviewed 60 respondents (including eight members of ethnic-minority groups), consisting of 41 patients, 16 healthcare professionals, one donor, one carer and one renal consultant's secretary. These respondents completed a DCE questionnaire and ranked potential priority criteria for renal transplantation. Most of the 60 respondents in the pilot exercise came from the University Hospitals of Coventry and Warwickshire (UHCW) NHS Trust, although 4 came from Ealing Hospital NHS Trust to boost responses (for full details refer to section 3.1 in chapter 5).

3.2. Selection of attributes and levels.

Details of the approach used to select final DCE attributes and levels are provided in section 3.1 of chapter 5. The final attributes and levels are listed in table 1 below (identical to that in chapter 5), reproduced here to remind the reader.

Table 1: Final attributes and levels.

Attribute	Variable name	Levels	Interpretation of coefficients.
Time spent awaiting transplantation	wait	1 month, 2 years, and 10 years.	Indirect utility of each 1 year reduction in transplant recipient waiting time.
Tissue type matching	tiss	Non-favourable match: 86% average kidney survival rate post-transplant. Favourable match: 89% average kidney survival rate post-transplant. Perfect match: 90% average kidney survival rate post-transplant.	Indirect utility of prioritizing people for each 1% improvement in kidney survival.
How many child or adult dependents recipients have	dep	None, 1, or 4 dependents.	Indirect utility of each additional dependent.
Recipient age	age	20 years, 45 years, and 65 years	Indirect utility for each 1 year reduction in recipient age.
Diseases predominantly affecting life expectancy	dis1	No disease affecting life expectancy (other than Kidney disease) vs. moderate disease (uncontrolled hypertension or obesity) & Kidney disease.	Indirect utility of having no rather than moderate disease predominantly affecting life expectancy.
	dis2	Moderate disease (uncontrolled hypertension or obesity) affecting life expectancy vs. severe disease (heart attack, stroke, or diabetes with complications).	Indirect utility of having moderate disease rather than severe disease predominantly affecting life expectancy.
Diseases predominantly affecting quality of life	ill1	No disease affecting quality of life (other than Kidney disease) vs. moderate disease (mild asthma).	Indirect utility of having no disease rather than a moderate disease predominantly affecting quality of life.
	ill2	Moderate disease (mild asthma) affecting quality of life vs. severe disease (severe arthritis).	Indirect utility of having a moderate disease rather than a severe disease predominantly affecting quality of life.

3.3. Design of the final questionnaire.

Details of the approach used to design the questionnaire can be found in section 3.3 of chapter 5. It should be noted that the questionnaire was made available in Punjabi, Hindi, Bengali, Gujarati and Urdu, as well as English. The questionnaires elicited information on gender and posed a question about ethnicity (see ethnicity question below). The first questionnaire in appendix E is the patient renal questionnaire in English.

Which of the following ethnic groups do you consider that you belong to?

(Please tick 1 box only):

White – British ❑

White – Irish ❑

White – Any other white background ❑
- please describe_____

Mixed – White / Black Caribbean ❑

Mixed – White / Black African ❑

Mixed – White / Asian ❑

Any other mixed background ❑
- please describe_____

Black or black British (Caribbean) ❑

Black or black British (African) ❑

Black or black British (Any other background ❑

Asian or Asian British (Indian) ❑

Asian or Asian British (Pakistani) ❑

Asian or Asian British (Bangladeshi) ❑

Asian or Asian British (Any other background) ❑

Chinese ❑

Any other ethnic group ❑
- please describe_____

3.4. Questionnaire distribution.

A total of 20 000 flyers with Freepost reply envelopes were enclosed in the UK National Kidney Federation's publication Kidney Life, inviting people to request questionnaires, including alternative-language versions if required. As we did not receive a large enough sample of ethnic-minority patients from the postal questionnaire, a bilingual researcher (Dr Anil Gumber) obtained 18 additional responses from members of ethnic minority groups at Ealing NHS Trust and five additional responses from members of ethnic-minority groups at University Hospitals of Coventry and Warwickshire NHS Trust.

3.5. Data analysis.

We used the same underlying econometric model to compare patient preferences for non-white ethnic-minority patients versus others (model 1), South Asian patients versus others (model 2), and female patients versus others (model 3). The term Y_{ij} is a binary dependent variable, from individuals i = 1...m, for observations j = 1...n_i. Observations n_i vary because the i individuals do not all complete every pairwise choice (some respondents do not answer all choices), μi is the random effects error term (which allows for multiple responses from i respondents), and ε_{ij} is the probit error term for individuals i for j observations.

Variables are defined in the Materials and Methods section. Ds is a dummy variable and is equal to 1 if the respondent is in the subgroup, otherwise it is equal to 0.

$$Y_{ij} = \beta_0+\beta_1 wait_{ij}+\beta_2 tiss_{ij}+\beta_3 dep_{ij}+ \beta_4 age_{ij}+\beta_5 dis1_{ij}+\beta_6 dis2_{ij}+\beta_7 ill1_{ij}+ \beta_8 ill2_{ij}+\beta_9 Ds_{ij}+\beta_{10} Ds_{ij} wait_{ij}+\beta_{11} Ds_{ij} tiss_{ij}+\beta_{12} Ds_{ij} dep_{ij}+\beta_{13} Ds_{ij} age_{ij}+\beta_{14} Ds_{ij} dis1_{ij}+\beta_{15} Ds_{ij} dis2_{ij}+ \beta_{16} Ds_{ij} ill1_{ij}+\beta_{17} Ds_{ij} ill2_{ij}+\mu_i+ \varepsilon_{ij}$$

(Models 1, 2, and 3).

3.6. Establishing the marginal rate of substitution (MRS).

MRS relates changes in attributes to a 1-year change in waiting time as a ratio. We used the Delta method (Wooldridge 2002) to establish whether MRS was significant. This was because the binary dependent variable model that we used (random effects probit) was non-linear, and the Delta method can be used to establish confidence intervals for estimated parameters for these types of models (Greene 2000). Moreover, the approach allows researchers to establish the significance or otherwise of a ratio of coefficients. Since MRS is a ratio, it allows clarification of whether MRS for a given variable is significant both for the defined subgroups of patients, and also for patients who are not in the defined ethnic specific or female subgroups (see Table 2). These tests for statistical significance were performed using the command 'nlcom' in STATA.

We also performed Wald tests to establish whether MRS in a subgroup differed in a statistically significant manner from MRS among other patients, in other words whether the non-white ethnic minorities, South Asian ethnic minorities or female subgroups had a different MRS to other patients who were not in that subgroup. So, for example, in relation to the variable tissue, the test we conducted was whether $\beta_2 / \beta_1 = (\beta_2 / \beta_{11})/(\beta_1 / \beta_{10})$.

These tests were performed using the command 'testnl' in STATA. Wald tests establish whether there is a significant difference in MRS comparing MRS for base groups, versus defined subgroups for each attribute. Differences in MRS at the 5% level are indicated by P-values of ≤0.05.

Table 2 Calculation of MRS

Variable	Base group MRS	Variable	MRS for subgroup of respondents
Waiting time	N /A	Waiting time	N /A
Tissue	β_2 / β_1	Tissue	$(\beta_2 + \beta_{11})/ (\beta_1 + \beta_{10})$
Dependent	β_3 / β_1	Dependent	$(\beta_3 + \beta_{12})/ (\beta_1 + \beta_{10})$
Age	β_4 / β_1	Age	$(\beta_4 + \beta_{13})/ (\beta_1 + \beta_{10})$
Disease1	β_5 / β_1	Disease1	$(\beta_5 + \beta_{14})/ (\beta_1 + \beta_{10})$
Disease2	β_6 / β_1	Disease2	$(\beta_6 + \beta_{15})/ (\beta_1 + \beta_{10})$
ill1	β_7 / β_1	ill1	$(\beta_7 + \beta_{16})/ (\beta_1 + \beta_{10})$
ill2	β_8 / β_1	ill2	$(\beta_8 + \beta_{17})/ (\beta_1 + \beta_{10})$

4. Results.

4.1. Sample characteristics.

The UK National Kidney Federation, which publishes Kidney Life, could not provide us with data that might allow us to assess the representativeness of our sample, so instead we used data from the UK Renal Registry (Farrington, Hodsman et al. 2008; Farrington, Udayaraj et al. 2008).

In total, 895 out of 908 respondents indicated their ethnic origin. Of these, 799 out of 895 patients (89.3%) were white (British), and 27 out of 908 (3%) were members of white ethnic minorities, so overall 92.3% of our sample was white. This compares with incidence data (Farrington, Udayaraj et al. 2008) which suggest that, across the UK, 79.8% of renal patients are white, so in our sample white patients were over-represented. Moreover, 69 out of 895 patients (7.7%) were members of non-white ethnic minorities, compared with a 17.9% incidence rate (Farrington, Udayaraj et al. 2008). Of the 69 members of nonwhite ethnic minorities, 50 patients were of South Asian origin. Therefore, 50 of the 895 patients in our sample (5.6%) were of South Asian origin, compared with a 10% incidence rate (Farrington, Udayaraj et al. 2008).

Members of non-white ethnic minorities consisted of two out of 69 mixed (white/black Caribbean), one out of 69 mixed (white/black African), one out of 69 mixed (white South Asian), two out of 69 with any other mixed background, including a Luso-Indian, one out of 69 Anglo-Indian / English-Portuguese. In total, seven out of 69 were black or black British Caribbean, three out of 69 were black or black British (African), one out of 69 was black or black British (any other background), and two patients were Chinese. The 50 South Asian patients in the non-white sample included 29 out of 69 South Asian or South Asian British (Indian) patients, nine out of 69 South Asian or South Asian British (Pakistani) patients, two out of 69 South Asian or South Asian British Bangladeshi patients, seven out of 69 South Asian or South Asian British (any other background) patients, plus one Filipina, one Persian and one Iranian patient. In total, 508 out of 908 patients (55.9%) were male, 397 out of 908 patients (43.7%) were female, and three out of 908 patients (0.3%) did not indicate their gender. This is reassuring, as Renal Registry data that have been presented graphically (Farrington, Udayaraj et al. 2008) show a trend towards slightly higher proportions of men than women among renal patients for all age groups.

The average patient age was 54.88 years (median 57 years). For members of white ethnic minorities the average age was 55.65 years (median 57 years), for those belonging to non-white ethnic minorities it was 54.12 years (median 56 years), and for patients of South Asian origin it was 55.38 years (median 56.5 years). Among male patients (508/908) the average age was 56.49 years (median 58 years), and among female patients (397/908) it was 52.85 years (median 54 years). Unfortunately the Renal Registry data (Farrington, Hodsman et al. 2008) are not specific for ethnic origin or gender. However, the median age for all patients is 56.9 years, which is remarkably close to our figure of 57 years.

The sample consisted of 468 out of 908 patients (51.5%) with successful transplants, 118 out of 908 patients (13%) whose transplant failed, and 279 out of 908 patients (30.7%)

who were awaiting transplants, with an average waiting period of 22.6 months. Some patients whose transplant failed are also included in the data for those awaiting transplants. This also applies to all gender and ethnic-minority groups. A total of 237 out of 908 patients (26.3%) were on dialysis without transplantation, and 57 out of 908 patients (6.3%) had kidney disease that did not require dialysis. Renal Registry prevalence data (Farrington, Hodsman et al. 2008) suggest that 46.6% of patients have successful transplants (as this is their current treatment modality), which is reassuringly close to our figure. However, there are no data for patients with failed transplants, or for those awaiting transplants, on dialysis without transplantation, or with kidney disease not requiring transplantation. Among non-white ethnic minorities there were 18 out of 69 patients (26%) with successful transplants, 10 out of 69 patients (14.5%) whose transplant failed, 35 out of 69 patients (50.7%) awaiting a transplant on dialysis (average waiting period 21.45 months), and three out of 69 patients (4.3%) with kidney disease not requiring dialysis. Among those of South Asian origin, 10 out of 50 patients (20%) had successful transplants, eight out of 50 patients (16%) had failed transplants, 28 out of 50 patients (56%) were awaiting transplants (average waiting period 23.1 months), and three out of 50 patients (6%) were on dialysis without transplantation. Unfortunately, the available data (Farrington, Hodsman et al. 2008) were not analysed by ethnic origin. However, given the shortage of transplants available to ethnic-minority groups, and their lower success rates, because they are likely to be poorer tissue matches, the lower percentage figure for transplant successes and the higher percentage figure for transplant failures might be expected.

4.2. Data analysis.

The results for models 1 to 3 are presented in Tables 3, 4 and 5.

4.2.1. Non-white ethnic minorities vs. other patients.

The likelihood ratio test for model 1 (see Table 3) is significant, which suggests that preferences do vary between members of non-white ethnic minorities and other patients.

Table 3 Model 1: patients – dummy variables for non-white ethnic minority patients

Attribute	Coefficient excluding non-white ethnic minorities	MRS excluding non-white ethnic minorities	Coefficient for dummy variables for non-white ethnic minorities	MRS for non-white ethnic minorities	Wald test (p-value)
wait	0.448**	1	-0.0025	1	
tiss	0.690**	1.54** (1.19 / 1.89)	-0.0718**	-0.07 (-1.13 / 1.00)	<0.001
dep	0.0605**	1.35** (1.08 / 1.62)	0.03450	2.26**(1.16 / 3.36)	0.311
age	0.0074**	0.16** (0.13 / 0.20)	-0.0045	0.07 (-0.05 / 0.19)	<0.001
dis1	0.0067	0.15 (-0.90 / 1.21)	-0.0773	-1.67 (-5.55 / 2.21)	0.375
dis2	0.7138**	15.93**(13.96/17.91)	-0.3649**	8.25**(2.86 /13.64)	<0.001
ill1	-0.1113**	-2.48** (-1.16 /-3.81)	-0.1049	-5.11*(-0.45 /-9.78)	0.992
ill2	0.1829**	4.08** (2.99 / 5.18)	0.0263	4.95* (0.97 / 8.92)	0.149
Intercepts	0.1306**		-0.0952		
Percentage of actual values predicted	62.64%	Sample	908 patients (69 are non-white ethnic minorities)	Mc Fadden's R^2	0.113
LR test (λ)	29.14	Dummy variables jointly significant	Yes CV for 9 dfs = 16.92	Log-likelihood	-4987.2

*Denotes significance at 1% level; **Denotes significance at 5% level but not at 1% level.

The Wald tests for three variables are also significant, which suggests that MRS differs significantly between the two patient groups for these three variables. For non-white ethnic minorities, MRS on the variable tiss is non-significant. This relates to prioritizing recipients with a good tissue match, so members of non-white ethnic minorities would not prioritize to recipients with better tissue matches. For other patients it is positive and significant, implying a preference for prioritising recipients with better tissue matches. Another difference relates to age. Among members of non-white ethnic minorities the variable age is non-significant, so they would not prioritise younger recipients, whereas

among other patients this variable is positive and significant, suggesting a preference for prioritizing younger recipients. Finally, there is evidence that preferences vary in relation to prioritizing those with diseases that affect life expectancy. The variable dis2 relates to prioritizing those with moderate rather than severe diseases that affect life expectancy. Members of non-white ethnic minorities place less emphasis than do other patients on prioritizing those with moderate rather than severe diseases that affect life expectancy (MRS = 8.25 vs. 15.93).

4.2.2. South Asian patients vs. other patients.

A similar pattern emerges in the South Asian patient sample (see Table 4), which is not unexpected, as they represented a large proportion (50 out of 69) of the non-white ethnic-minority group. Once again likelihood ratio tests suggest that preferences do vary between the two patient groups, and the Wald tests suggest that these differences relate to the same three variables. There is no evidence that South Asian patients would prioritize those with a better tissue match, as the variable tiss is non-significant. However, among other patients, the variable is positive and significant, which suggests a preference for prioritizing recipients with better tissue matches. South Asian patients would not prioritize the young rather than the old, as the variable age is non-significant, whereas among other patients it is positive and significant. Finally, although both South Asian patients and the rest of the patient sample would prioritize those with moderate (dis2) rather than severe diseases that affect life expectancy, South Asian patients would be less likely to prioritize on the basis of this criterion (MRS = 7.57 vs. 15.78).

Table 4 Model 2: patients – dummy variables for South Asian patients

Attribute	Coefficient excluding South Asian ethnic minorities	MRS excluding South Asian ethnic minorities	Coefficient for South Asian ethnic minorities	MRS for South Asian ethnic minorities	Wald test (p-value)
wait	0.0450**	1	-0.0069	1	
tiss	0.0681**	1.51** (1.17/1.85)	-0.0824**	-0.38 (-1.74 / 0.99)	0.001
dep	0.0609**	1.36** (1.09 /1.62)	0.0386	2.61** (1.07 / 4.16)	0.434
age	0.0073**	0.16** (0.13 / 0.20)	-0.0055	0.048 (-0.11 / 0.21)	0.002
dis1	0.0023	0.05 (-0.99/1.08)	-0.0243	-0.58 (-5.58 / 4.43)	0.803
dis2	0.7095**	15.78**(13.84/17.71)	-0.4214**	7.57**(0.71 / 14.42)	<0.001
ill1	-0.1119**	-2.49**(-1.19 / -3.80)	-0.1371	-6.54*(-0.42/-12.65)	0.827
ill2	0.1807**	4.02** (2.94 / 5.10)	0.0785	6.81* (1.49 / 8.92)	0.417
Intercepts	0.1286**		-0.0932		
Percentage of actual values predicted	62.71%	Sample	908 patients (50 are South Asian ethnic minorities)	Mc Fadden's R^2	0.113
LR test (λ)	27.76	Dummy variables jointly significant	Yes CV for 9 dfs = 16.92	Log-likelihood	-4987.85

*Denotes significance at 1% level; **Denotes significance at 5% level but not at 1% level.

4.2.3. Preferences and gender.

The results of the likelihood ratio test do not provide evidence of a difference in preferences between male and female patients (see Table 5). However, Wald tests suggest that preferences may vary in relation to four out of eight variables. These tests suggest that preferences vary in relation to prioritizing on the basis of tissue match (tiss). Both male and female patients valued this criterion significantly. However, it appears that females value it marginally more than do males (MRS = 1.45 vs. 1.34). The Wald test also suggests that preferences differ with regard to prioritizing recipients with child or adult dependants. The variable 'dep' is significant for both groups, but female patients appear to value this marginally more (MRS = 1.61 vs.1.28). The Wald test suggests that preferences for prioritizing younger rather than older dependents might also differ. Female patients place marginally more emphasis on this variable (age) than do males (MRS = 0.17 vs.

0.14). Finally, both female and male patients value prioritizing those with severe rather than moderate diseases that affect life expectancy (dise2) significantly. However, this variable seems to be valued marginally less by female patients (MRS = 14.86 vs. 15.43).

Table 5 Model 3: patients with female patient dummy variables

	Coefficient for male patients	MRS for male patients	Coefficient for female patients	MRS for female patients	Wald test (p-value)
wait	0.0448**	1	-0.0003	1	
tiss	0.0603**	1.34** (0.90 / 1.78)	0.0045	1.45** (0.95 / 1.96)	0.009
dep	0.0575**	1.28** (0.94 / 1.62)	0.0141	1.61** (1.20 / 2.01)	0.014
age	0.0064**	0.14** (0.10 / 0.19)	0.0011	0.17** (0.11 / 0.22)	0.026
dis1	-0.0373	-0.83 (-2.21 / 0.54)	0.0704	0.74 (-0.80 / 2.28)	0.137
dis2	0.6917**	15.43**(12.91/ 17.93)	-0.0295	14.86**(12.07/17.64)	<0.001
ill1	-0.1150**	-2.56** (-0.85 /-4.27)	-0.0131	-2.87** (-0.94 /-4.80)	0.285
ill2	0.1615**	3.60** (2.19 / 5.01)	0.0520	4.79** (3.18 / 6.40)	0.175
Intercepts	0.1144**		0.0201		
Percentage of actual values predicted	62.50%	Sample	908 patients (397 are female)	Mc Fadden's R^2	0.110
LR test (λ)	5.00	Dummy variables jointly significant	Yes CV for 9 dfs = 16.92	Log-likelihood	-4908.40

*Denotes significance at 1% level; **Denotes significance at 5% level but not at 1% level.

4.2.4. Summary of how preferences differ by ethnicity and gender.

These findings suggest that patients who are not members of ethnic minorities value prioritizing patients with closer tissue matches, whereas South Asian patients and those from non-white ethnic minorities do not. Patients in general, including those who belong to ethnic minorities, prioritize those who have had to wait a long time for a transplant, and those with child or adult dependents. However, prioritizing younger people is not valued among South Asians and non-white ethnic minorities, whereas it is among other patients. Those with moderate, rather than severe diseases that affect life expectancy are a priority for patients in general, but less of a priority among South Asian patients and non-white

ethnic minorities. All ethnic groups value prioritizing those with moderate, as opposed to no, disease that affect quality of life. This may seem a somewhat odd result, but it could be explained by enlightened self-interest, in that many respondents themselves would have moderate disease in addition to kidney disease, which affect their quality of life. Moreover, there is no evidence that the ethnic-minority groups value prioritizing those with moderate rather than severe diseases that affect quality of life differently. Both groups would prioritize potential recipients with moderate rather than severe diseases that affect quality of life.

Although there is evidence that preferences vary according to gender, these differences are not particularly pronounced. However, women do have a slightly greater tendency to prioritize recipients who are better tissue matches to donors. Women are also slightly more likely to prioritize those with child or adult dependents, and younger people, and slightly less likely to prioritize those with moderate rather than severe diseases that affect life expectancy.

5. Discussion.

Discrete choice experiments (DCEs) are increasingly used in health technology assessment (Ryan 1999) and health economics (Ryan and Gerard 2003). Indeed, searches on PubMed have identified several hundred health-related DCEs. However, although some DCEs have addressed the concerns of ethnic minorities (Bennett and Savani 2004; Dwight-Johnson, Lagomasino et al. 2004; Byrne, Souchek et al. 2006; Hall, Fiebig et al. 2006; Peacock, Apicella et al. 2006; Hawley, Volk et al. 2008; Lee, Brooks et al. 2008; Constantinescu, Goucher et al. 2009; Bridges, Selck et al. 2011; Thrumurthy, Morris et al. 2011), the majority have assessed preferences for respondents overall, rather than distinguishing those of minority groups. Also a small number of DCEs have looked at whether preferences vary according to gender (Brown, Swinyard et al. 2003; Mays and

Zimet 2004; Tsang, Chan et al. 2004; Kjaer, Gyrd-Hansen et al. 2006; Hjelmgren and Anell 2007; Gerard, Salisbury et al. 2008; Goto, Takahashi et al. 2009; Hauber, Mohamed et al. 2009; Mentzakis, Stefanowska et al. 2011; Thrumurthy, Morris et al. 2011). This chapter has therefore addressed both these shortcomings in the context of a DCE used to evaluate preferences for kidney allocation.

DCEs have strong theoretical foundations in economics. They are compatible with Lancaster's characteristics theory of demand (Lancaster 1966) and random utility theory (McFadden 1999). They are often used to establish how much people are willing to pay for different attributes of healthcare provision. However, there are methodological issues which need to be addressed before it can be assumed that DCE estimates of willingness to pay (WTP) are accurate (Ryan, McIntosh et al. 1998; Ratcliffe 2000; Ryan and Farrar 2000; Ryan, Watson et al. 2003). One major concern is that, if they are applied in a context in which healthcare is free at the point of use (see chapters 3 and 4), respondents may indicate an unrealistically high WTP because they know that they will not in fact bear a cost, leading to hypothetical bias.

We did not elicit WTP in this DCE, thereby avoiding many of these potential problems. However, it must be conceded that our results are sensitive to the choice of attributes selected, and can only give an indication of trade-offs in relation to the actual attributes included. Since there are no definitive criteria for establishing the appropriate attributes and levels to include in a DCE, researchers simply have to consult a wide range of opinion, including patients and professionals, before deciding upon which attributes and levels to include, and ensure that their choice of attributes has emerged from a thorough pilot exercise. This is why we invested a great deal of time piloting the questionnaire.

Although DCEs have been applied to determine priorities for UK liver transplants (Ratcliffe and Buxton 1999; Ratcliffe 2000), that particular study did not collect ethnicity data, only

gender data. Also, although the study reported differences in responses by gender, the data were not analyzed to establish whether preferences varied with gender. The only other DCE work in the area of transplantation available in 2009 when we published our paper (Clark, Gumber et al. 2009) was another UK study of factors that influence people's willingness to donate body parts for transplantation in the event of their death (Bennett and Savani 2004). This considered three groups (white, South Asian and Afro-Caribbean) but concluded that 'being of a particular ethnicity or gender did not affect outcomes in any meaningful ways', so the authors only reported results for respondents overall (Bennett and Savani 2004). More recently, however, a Canadian analysis has been published (Davison, Kromm et al. 2010) which applies DCEs to assess patient and professional preferences for organ allocation and procurement, end-of-life care, and organization of care for patients with chronic kidney disease. Information on respondents' race and gender were collated but, once again, the issue of whether preferences varied by ethnicity and gender was not addressed. More to the point the DCE only contained one attribute relating to kidney allocation policy which was 'How should deceased donor kidneys be allocated for transplantation' which had levels of 'first come first served' or 'best match.' In contrast to the renal transplant DCE presented in this thesis, most of the DCE's attributes concerned themselves with other issues such as who should provide dialysis; how should kidneys be obtained; when should end of life discussions be started; how much information on end of life care issues should be provided; and how should decisions to stop dialysis be made.

In the field of transplantation there are of course other studies which do not use DCE methodology. Such kidney allocation studies have been conducted in Australia and America (Louis, Sankar et al. 1997; Browning and Thomas 2001), and may not be generalizable to the UK. The Australian study (Browning and Thomas 2001) involved respondents ranking possible priority criteria for transplantation, including age, gender, occupation, education, work status, income, whether potential recipients were parents,

post-transplantation prognosis, and length of time for which recipients had been on the transplant list. They therefore avoided addressing the issue of whether to prioritize on the basis of ethnicity. They found that over 90% of 238 respondents considered that recipient gender, socioeconomic status, employment status and occupation should not influence decisions about kidney transplant allocation. Instead, most of the respondents (87.4%) considered that those who had been on the transplant list for a long period of time should have priority, and 79% would prioritize those with a good prognosis, whilst 65% would prioritize younger recipients.

The American study (Louis, Sankar et al. 1997) noted that the American point-based allocation system disadvantaged African Americans because of its emphasis on antigen matching, as African Americans typically have a disproportionate number of rare antigens. They used semi-structured interviews with 33 patients who were awaiting transplants, including some black Americans, who considered that discrimination in organ allocation by antigen matching was unfair. However, there was a paradox in that they did not want to receive organs that gave them a reduced likelihood of survival. So these results differ from ours, but of course the rate of graft survival has increased since the American study because of improvements in anti-rejection drugs, so this may partly explain the differences in findings, as may the small sample and methodology used. The authors did not address the issue of gender-related differences.

There is one other study (Geddes, Rodger et al. 2005) which was conducted in Scotland. A total of 295 respondents were asked to choose one of two hypothetical patients from eight scenarios to establish whether the patients agreed with the current criteria for transplant allocation in the UK. Ethnicity was not taken into consideration in this research, although gender was addressed. The findings suggested that neither age nor gender of the recipient should be used when making decisions about the allocation of kidneys. The former is somewhat at odds with our findings for the white majority patients who, unlike

the ethnic minorities, would tend to prioritize younger recipients. This research was conducted prior to the UK Transplant 2006 reforms to transplant allocation criteria. It seemed to broadly support a shift away from the previous emphasis on tissue matching. The main conclusion was that allocation should favour respondents who had waited for longer, and of course UK transplant policy did evolve to place more emphasis on those who have waited a long time for a transplant.

6. Conclusions.

Our main conclusions, in relation to the methodology used in this chapter, are that it appeared to perform well in terms of demonstrating that the non-white ethnic minorities and Asian ethnic minorities have distinctly different preferences in relation to priority criteria for renal transplantation, compared with patients who are not in these minority groups. The use of interaction dummy variables also allowed us to demonstrate that preferences for specific attributes differed in a statistically significant manner. In addition the approach we adopted of expressing results in terms of Marginal Rates of Substitution (MRS) with respect to waiting time worked well. Moreover, using Wald tests to establish whether MRS for attributes differed by ethnicity served to further reinforce our findings.

In relation to whether preferences varied according to the gender of respondents, we obtained results which were conflicting. The individual interaction dummy variables for female patients were all insignificant. However, Wald test results which looked at whether MRS for an attribute varies between the male and female respondent group, suggested that there was evidence that MRS varied in a statistically significant manner for 4 variables (even though actual differences in MRS were generally not that large).

Our DCE findings clearly have policy implications in that they are broadly supportive of revisions to UK transplant kidney allocation policy in 2006, which reduced the emphasis

on transplanting to patients with good tissue matches. However, although the policy shift placed less emphasis on tissue matching as an allocation criterion, current policy still retains quality of tissue matching as an allocation criterion. Even though this might be supported by the majority of patients, evidence from the research presented in this chapter suggests that it would not be supported by South Asians and members of non-white ethnic minorities more generally. Non-white ethnic minorities and South Asians would prefer the quality of tissue type matching between donor and recipient to be abandoned as a criterion for allocation. They are disadvantaged if transplant allocation is based on tissue matching, which no doubt accounts for this finding. UK Transplant's policy shift towards prioritizing those who have waited a long time for a transplant is supported by our findings for all ethnic-minority groups, irrespective of gender. However, the other shift in emphasis, towards prioritizing younger patients, does not appear to be supported by ethnic-minority groups, although it is supported by other patients.

Finally, although we have found some evidence that preferences do vary with gender, these differences are not particularly pronounced, which suggests that an attempt to facilitate the preferences of people according to gender is a low priority, and that addressing the specific needs and disadvantages of ethnic-minority groups should be a more urgent consideration, when transplant policy is reassessed.

7. References.

Bellary, S., J. P. O'Hare, et al. (2008). "Enhanced diabetes care to patients of south Asian ethnic origin (the United Kingdom Asian Diabetes Study): a cluster randomised controlled trial." Lancet **371**(9626): 1769-1776.

Bennett, R. and S. Savani (2004). "Factors influencing the willingness to donate body parts for transplantation." J Health Soc Policy **18**(3): 61-85.

Bridges, J. F. P., F. W. Selck, et al. (2011). "Condom avoidance and determinants of demand for male circumcision in Johannesburg, South Africa." Health Policy and Planning **26**(4): 298-306.

Brown, A. J., W. Swinyard, et al. (2003). "Women in academic medicine: A report of focus groups and questionnaires, with conjoint analysis." Journal of Womens Health **12**(10): 999-1008.

Browning, C. J. and S. A. Thomas (2001). "Community values and preferences in transplantation organ allocation decisions." Soc Sci Med **52**(6): 853-861.

Byrne, M. M., J. Souchek, et al. (2006). "Racial/ethnic differences in preferences for total knee replacement surgery." Journal of Clinical Epidemiology **59**(10): 1078-1086.

Churak, J. M. (2005). "Racial and ethnic disparities in renal transplantation." J Natl Med Assoc **97**(2): 153-160.

Clark, M. D., A. K. Gumber, et al. (2009). "Prioritising patients for renal transplantation? Analysis of patient preferences for kidney allocation according to ethnicity and gender." Diversity in Health and Care, 2009 **6**: 181-191.

Constantinescu, F., S. Goucher, et al. (2009). "Understanding Why Rheumatoid Arthritis Patient Treatment Preferences Differ by Race." Arthritis & Rheumatism-Arthritis Care & Research **61**(4): 413-418.

Davis, C. and G. Randhawa (2006). "The influence of religion on organ donation and transplantation among the Black Caribbean and Black African population - A pilot study in the United Kingdom." Ethnicity & Disease **16**(1): 281-285.

Davison, S. N., S. K. Kromm, et al. (2010). "Patient and health professional preferences for organ allocation and procurement, end-of-life care and organization of care for patients with chronic kidney disease using a discrete choice experiment." Nephrol Dial Transplant **25**(7): 2334-2341.

Department of Health (2008). "Organs for transplants: a report from the Organ Donation Taskforce." [http--www.dh.gov.uk-prod_consum_dh-groups-dh_digitalassets-@dh-@en-documents-digitalasset-dh_082120.pdf](16/01/2008).

Dwight-Johnson, M., I. T. Lagomasino, et al. (2004). "Using conjoint analysis to assess depression treatment preferences among low-income Latinos." Psychiatric Services **55**(8): 934-936.

Essers, B. A., C. D. Dirksen, et al. (2010). "Assessing the public's preference for surgical treatment of primary basal cell carcinoma: a discrete-choice experiment in the south of the Netherlands." Dermatol Surg **36**(12): 1950-1955.

Faggioli, G., L. Scalone, et al. (2011). "Preferences of patients, their family caregivers and vascular surgeons in the choice of abdominal aortic aneurysms treatment options: the PREFER study." Eur J Vasc Endovasc Surg **42**(1): 26-34.

Farrington, K., A. Hodsman, et al. (2008). "ERSD prevalent rates in 2007 in the UK: national and centre specific analyses." In: UK Renal Registry 2008 Report; www.renalreg.org (accessed April 2009).

Farrington, K., U. Udayaraj, et al. (2008). "ESRD incident rates in 2007 in the UK: national and centre-specific analyses." In UK Renal Registry 2008 Report: www.renalreg.org (accessed April 2009).

Geddes, C. C., R. S. Rodger, et al. (2005). "Allocation of deceased donor kidneys for transplantation: opinions of patients with CKD." American Journal of Kidney Diseases **46**(5): 949-956.

Gerard, K., C. Salisbury, et al. (2008). "Is fast access to general practice all that should matter? A discrete choice experiment of patients' preferences." J Health Serv Res Policy **13**: 3-10.

Goto, R., Y. Takahashi, et al. (2009). "A cohort study to examine whether time and risk preference is related to smoking cessation success." Addiction **104**(6): 1018-1024.

Greene, W. H. (2000). "Functional form, nonlinearity and specification." In Econometric Analysis 4th edn.(Upper Saddle River NJ: Prentice Hall International Incorporated).

Guimaraes, C., C. A. Marra, et al. (2009). "A valuation of patients' willingness-to-pay for insulin delivery in diabetes." Int J Technol Assess Health Care **25**(3): 359-366.

Hall, J., D. G. Fiebig, et al. (2006). "What influences participation in genetic carrier testing? Results from a discrete choice experiment." J Health Econ **25**(3): 520-537.

Hauber, A. B., A. F. Mohamed, et al. (2009). "Educational and psychological aspects of treatment preferences and medication adherence of people with type 2 diabetes using oral glucose-lowering agents." Diabetic Medicine **26**: 416-424.

Hawley, S. T., R. J. Volk, et al. (2008). "Preferences for colorectal cancer screening among racially/ethnically diverse primary care patients." Med Care **46**(9 Suppl 1): S10-16.

Higgins, R. M., N. West, et al. (1997). "Effect of a strict HLA matching policy on distribution of cadaveric kidney transplants to Indo-Asian and white European recipients: regional study." BMJ **315**(7119): 1354-1355.

Hjelmgren, J. and A. Anell (2007). "Population preferences and choice of primary care models: A discrete choice experiment in Sweden." Health Policy **83**(2-3): 314-322.

Kjaer, T., D. Gyrd-Hansen, et al. (2006). "Investigating patients' preferences for cardiac rehabilitation in Denmark." Int J Technol Assess Health Care **22**(2): 211-218.

Koene, R. A. P. (2002). "Should the allocation of cadaveric kidneys for transplantation be based on HLA matching?" Nephrology Dialysis Transplantation **17**(5): 717-718.

Lancaster, K. J. (1966). "New Approach to Consumer Theory." Journal of Political Economy **74**(2): 132-157.

Lee, S. J., R. A. Brooks, et al. (2008). "HIV vaccine acceptability among immigrant Thai residents in Los Angeles: a mixed-method approach." Aids Care-Psychological and Socio-Medical Aspects of Aids/Hiv **20**(10): 1161-1168.

Louis, O. N., P. Sankar, et al. (1997). "Kidney transplant candidates' views of the transplant allocation system." J Gen Intern Med **12**(8): 478-484.

Mays, R. M. and G. D. Zimet (2004). "Recommending STI vaccination to parents of adolescents: The attitudes of nurse practitioners." Sexually Transmitted Diseases **31**(7): 428-432.

McFadden, D. (1999). "Computing willingness-to-pay in random utility models " In: Melvin JR, Moore JC and Riezman R (eds) Trade, Theory and Econometrics: essays in honor of John S Chipman. : 253–274.

Mentzakis, E., P. Stefanowska, et al. (2011). "A discrete choice experiment investigating preferences for funding drugs used to treat orphan diseases: an exploratory study." Health Econ Policy Law **6**(3): 405-433.

Peacock, S., C. Apicella, et al. (2006). "A discrete choice experiment of preferences for genetic counselling among Jewish women seeking cancer genetics services." Br J Cancer **95**(10): 1448-1453.

Pereira, C. C., M. Mulligan, et al. (2011). "Determinants of influenza vaccine purchasing decision in the US: a conjoint analysis." Vaccine **29**(7): 1443-1447.

Raleigh, V. S. (1997). "Diabetes and hypertension in Britain's ethnic minorities: implications for the future of renal services." BMJ **314**(7075): 209-213.

Ratcliffe, J. (2000). "Public preferences for the allocation of donor liver grafts for transplantation." Health Econ **9**(2): 137-148.

Ratcliffe, J. (2000). "The use of conjoint analysis to elicit willingness-to-pay values. Proceed with caution?" Int J Technol Assess Health Care **16**(1): 270-275.

Ratcliffe, J. and M. Buxton (1999). "Patients' preferences regarding the process and outcomes of life-saving technology. An application of conjoint analysis to liver transplantation." Int J Technol Assess Health Care **15**(2): 340-351.

Ryan, M. (1999). "A role for conjoint analysis in technology assessment in health care?" Int J Technol Assess Health Care **15**(3): 443-457.

Ryan, M. and S. Farrar (2000). "Using conjoint analysis to elicit preferences for health care." British Medical Journal **320**(7248): 1530-1533.

Ryan, M. and K. Gerard (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., E. McIntosh, et al. (1998). "Methodological issues in the application of conjoint analysis in health care." Health Econ **7**(4): 373-378.

Ryan, M., V. Watson, et al. (2003). "Methodological issues in the monetary valuation of benefits in healthcare." Expert Rev Pharmacoecon Outcomes Res **3**(6): 717-727.

Scalone, L., L. G. Mantovani, et al. (2009). "Patients', physicians', and pharmacists' preferences towards coagulation factor concentrates to treat haemophilia with inhibitors: results from the COHIBA Study." Haemophilia **15**(2): 473-486.

Thrumurthy, S. G., J. J. Morris, et al. (2011). "Discrete-choice preference comparison between patients and doctors for the surgical management of oesophagogastric cancer." Br J Surg **98**(8): 1124-1131; discussion 1132.

Tinelli, M., M. Ryan, et al. (2009). "Patients' preferences for an increased pharmacist role in the management of drug therapy." Int J Pharm Pract **17**(5): 275-282.

Torbica, A. and G. Fattore (2010). "Understanding the impact of economic evidence on clinical decision making: a discrete choice experiment in cardiology." Soc Sci Med **70**(10): 1536-1543.

Tsang, H. W. H., F. Chan, et al. (2004). "Factors influencing occupational therapy students' attitudes toward persons with disabilities: A conjoint analysis." American Journal of Occupational Therapy **58**(4): 426-434.

UK Transplant (2006). "Transplant Activity in the UK. Bristol: Statistics and Audit Directorate."

van Dam, L., L. Hol, et al. (2010). "What determines individuals' preferences for colorectal cancer screening programmes? A discrete choice experiment." European Journal of Cancer **46**(1): 150-159.

van Til, J. A., A. M. Stiggelbout, et al. (2009). "The effect of information on preferences stated in a choice-based conjoint analysis." Patient Educ Couns **74**(2): 264-271.

Wooldridge, J. M. (2002). "Economic analysis of cross section and panel data." MIT press.

Chapter 7: When simple may be more efficient - Econometric modelling of patient discrete choice experiment (DCE) data. Exploring preference heterogeneity, using Mixed Logit, a Latent Class Model, or Conditional Logit with dummy variables.

1. Introduction.

In Chapters 5 and 6 issues relating to differences in preferences between respondents' groups for the renal transplant project were addressed using interaction dummy variables. In this chapter I explore the application of both Mixed Logit and Latent Class Modelling (LCM) which expose unobserved preference heterogeneity, upon data relating to patient preferences for renal transplants. As indicated in Chapter 1 (section D), Mixed Logit is now increasingly used to establish DCE preferences econometrically in a manner that, at the same time, facilitates some assessment of whether preferences are heterogeneous. In addition, Latent Class Models (LCM) have also begun to be used in recent years. Such models provide an alternative form of econometric analysis which can be used to assess whether DCE respondents' preferences are characterized by preference heterogeneity at all. Prior to the introduction of these econometric approaches, health economists often relied on the application of interaction dummy variables to explore observable preference heterogeneity using DCE data. In this chapter, I compare all 3 approaches to examining preference heterogeneity, alongside models which do not cater for heterogeneity. This is to establish what additional information these methods catering for preference heterogeneity provide. The techniques are applied to patient DCE data relating to respondents' preferences for alternative priority criteria for kidney transplantation.

We analysed discrete choice experiment (DCE) data using Random Effects Logit, Conditional Logit, Mixed Logit, LCM, and Conditional Logit with dummy variables. Data were analysed for

a sample of 863 renal patient responses elicited using the patient renal transplant DCE questionnaire (whose responses were complete enough for them to be analysed using all 5 econometric models).

I recognize that both Mixed Logit and LCM are welcome additions to the health economist's tool box. This is because models which don't cater for preference heterogeneity and present average valuations might mask preference heterogeneity which Mixed Logit / LCM could expose. In addition, I wanted to compare the application of Mixed Logit and LCM to a more basic Conditional Logit model with interaction dummy variables. The reason for this is because I wanted to establish whether, if differences in preferences are anticipated for defined respondent groups, a simpler econometric model informed by prior expectations might be more revealing than reliance upon Mixed Logit / LCM (without the use of interaction dummy variables, or having an LCM which looks at the impact of respondent characteristics upon latent class membership).

2. Background.

Recently a number of discrete choice experiment analyses have been conducted which also assess preference heterogeneity using either Mixed Logit of Latent Class Models.

During 2001 – 2008 there were 6 such studies (Hall, Fiebig et al. 2006, Goto, Nishimura et al. 2007, Lancsar, Hall et al. 2007, Hole 2008, Kjae and Gyrd-Hansen 2008, Negrin, Pinilla et al. 2008) which used Mixed Logit (de Bekker-Grob, Ryan et al. 2012). One of these studies (Hole 2008) also conducted LCM alongside Mixed Logit (the only paper using such analysis during 2001 – 2008).

However, more recently for the period 2009 – 2011 there have been far more published analyses using Mixed Logit (Eberth, Watson et al. 2009, Hauber, Mohamed et al. 2009, Howard and Salkeld 2009, Ozdemir, Johnson et al. 2009, Regier, Friedman et al. 2009, van Helvoort-Postulart, Dellaert et al. 2009, van Helvoort-Postulart, van der Weijden et al. 2009, Blaauw, Erasmus et al. 2010, de Bekker-Grob, Hofman et al. 2010, Johnson, Ozdemir et al. 2010, Scuffham, Whitty et al. 2010, Wittink, Cary et al. 2010, Goto, Takahashi et al. 2011, Mohamed, Epstein et al. 2011, Oteng, Marra et al. 2011, Potoglou, Burge et al. 2011, Scalone, Watson et al. 2011, Sweeting, Whitty et al. 2011, Whitty, Scuffham et al. 2011) Also, during this period 3 analyses used LCM (Grindrod, Marra et al. 2010, Mentzakis, Ryan et al. 2011, Mentzakis, Stefanowska et al. 2011).

3. Methods.

The DCE study involves respondents making choices, about which one of two hypothetical transplant recipients (differing in characteristics) should receive a kidney. The data used is DCE patient data from the renal transplant study (full details are provided in Chapter 5). For the purposes of the analysis contained in this chapter, only the DCE responses obtained from patients are evaluated.

3.1. Pilot exercise.

Details of the thorough DCE pilot exercise conducted are provided in Chapter 5 (section 3.1).

3.2. Attributes and levels –final DCE.

Details of how attributes and levels were selected for the final DCE are provided in section 3.2 of Chapter 5. Details of final attributes and levels are in Table 1 below.

Table 1: Final attributes and levels.

Attribute	Variable name	Levels	Interpretation of coefficients.
Time spent awaiting transplantation	wait	1 month, 2 years, and 10 years.	Indirect utility of each 1 year reduction in transplant recipient waiting time.
Tissue type matching.	tiss	Non-favourable match: 86% average kidney survival rate post-transplant. Favourable match: 89% average kidney survival rate post-transplant. Perfect match: 90% average kidney survival rate post-transplant.	Indirect utility of prioritizing people for each 1% improvement in kidney survival.
How many child or adult dependents recipients have	dep	None, 1, or 4 dependents.	Indirect utility of each additional dependent.
Recipient age	age	20 years, 45 years, and 65 years	Indirect utility for each 1 year reduction in recipient age.
Diseases predominantly affecting life expectancy	dis1	Moderate disease (uncontrolled hypertension or obesity) & Kidney disease rather than no disease affecting life expectancy (other than Kidney disease).	Indirect utility of having moderate rather than no disease predominantly affecting life expectancy.
	dis2	Severe disease (heart attack, stroke, or diabetes with complications) rather than no disease affecting life expectancy (other than Kidney disease).	Indirect utility of having severe disease rather than no disease predominantly affecting life expectancy.
Diseases predominantly affecting quality of life	ill1	Moderate disease (mild asthma) rather than no disease affecting quality of life (other than Kidney disease)	Indirect utility of having moderate disease rather than no disease predominantly affecting quality of life.
	ill2	Severe disease (severe arthritis) affecting quality of life rather than no disease affecting quality of life (other than Kidney disease)	Indirect utility of having severe disease rather than no disease predominantly affecting quality of life.

3.3. Development of final DCE.

Details of design of the DCE are reported in Chapter 5 (section 3.3).

3.4. Questionnaire distribution.

Details of questionnaires distribution is provided in Chapter 5 (section 3.4). The patient questionnaire is the first one in appendix E.

3.5.0 Selection of Econometric models.

Some authors have argued that for forced choice DCE designs, random effects probit or random effects logit should normally be considered the econometric model of choice (Amaya-Amaya and Gerard 2008, de Bekker-Grob, Ryan et al. 2012). Such models automatically cater for multiple responses from each respondent..The increased interest in the use of conditional logit as an alternative to random effects logit or random effects probit is in part related to the increased use of DCE models with more than 2 choices. For example, when there are 2 choices plus an opt out (Kruijshaar, Essink-Bot et al. 2009), or where there are 3 choices (de Bekker-Grob, Essink-Bot et al. 2008), or 2 choices plus a do-nothing option (Rennie, Porteous et al. 2012)**.**

However, recently an increasing number of published DCEs have involved the application of conditional logit when only 2 options are compared without an opt-out (Guimaraes, Marra et al. 2009, Guimaraes, Marra et al. 2010, Stolk, Oppe et al. 2010, Swinburn, Lloyd et al. 2011, Park, Jo et al. 2012, Robyn, Barnighausen et al. 2012, McNamara, Chen et al. 2013). The use of conditional logit may be justified in such circumstances if conditional logit models are associated with improvements in the 'goodness of fit' in the model over and above models such as random effects logit. This is why in this analysis a basic random effects logit model

(3.5.1) is compared with a basic conditional logit model (3.5.2) to see which model appears to be associated with a better 'goodness of fit'.

I also wanted to apply conditional logit, because a comparison between conditional logit with interaction dummy variables alongside mixed logit or LCM, may then be generalizable to cases with more than 2 response options (for which random effects logit or random effects probit are not applicable). However, since the basic conditional logit model does not involve an error term catering for multiple responses from each respondent (like random effects logit or random effects problit) the error term needs correcting to allow for multiple responses from each respondent through the specification of a data clustering command. This chapter begins by applying two very basic models which do not provide any information relating to preference heterogeneity. They are random effects logit - model 1 (3.5.1) and conditional logit with a clustering error term controlling for multiple responses - model 2 (3.5.2). Thereafter, models which allow for unobserved preference heterogeneity are adopted including Mixed Logit - model 3 (3.5.3) and a Latent Class Model (LCM) -model 4 (3.5.4). Finally a conditional logit model with interaction dummy variables – model 5 (3.5.5) for ethnic minority respondents is adopted (including the clustering command). This model makes some preference heterogeneity not observable in models 1 and 2 (relating to differences in preferences amongst ethnic minority respondents) observable. However, it does not allow for assessment of other unobserved preference heterogeneity which the Mixed Logit and LCM models cater for. For the final analysis – model 5 (3.5.5) I decided I would apply a conditional logit model rather than a random effects model, if data analysis comparing model 1 (random effects logit) and 2 (conditional logit with an error correction term) showed that model 2 was associated with better 'goodness of fit' compared to model 1(measured by Mc Faddens R^2 and the % of values accurately predicted by the model).

3.5.1 Econometric model – Random Effects Logit.

All the econometric analysis was conducted in STATA. We first used a Random Effects Logit model (model 1), to establish patient preferences. Variables are defined in table 1. Here, we assume a utility model (U_{it}) which can be modeled as a function of the levels of attributes, using a binary dependent model (Y_{it}), for individuals (i = 1 to n), where 'n' denotes the total number of respondents, in each choice set (t = 1 to16). The term μ_i is the random effects error term and ξ_{it} is the other error term.

$$U_{it} = \beta_0 + \beta_1 wait_{it} + \beta_2 tiss_{it} + \beta_3 dep_{it} + \beta_4 age_{it} + \beta_5 dis1_{it} + \beta_6 dis2_{it} + \beta_7 ill1_{it} + \beta_8 ill2_{it} + \mu_i + \xi_{it}$$

$$Y_{it} = \beta_0 + \beta_1 wait_{it} + \beta_2 tiss_{it} + \beta_3 dep_{it} + \beta_4 age_{it} + \beta_5 dis1_{it} + \beta_6 dis2_{it} + \beta_7 ill1_{it} + \beta_8 ill2_{it} + \mu_i + \xi_{it}$$

(Model 1)

3.5.2. Econometric model – Conditional Logit.

Here, we again assume a utility model (U_{ijt}) which can be modeled using the dependent variable Y_{ijt} which is a function of the specified attribute variables, and ξ_{ijt} (the error term). The term μ_i caters for multiple responses from each respondent using the clustering command in STATA. With the conditional logit model (model 2) choice amongst alternatives is modeled as function of the characteristics of the alternatives, so it is well suited to estimating behavioural models (Hoffman and Duncan 1988). Here we have individuals (i = 1 to n), were 'n' denotes the total number of respondents who choose an alternative (j = 1 or 2), in each choice set (t = 1 to 16).

$$U_{ijt} = \beta_1 wait_{ijt} + \beta_2 tiss_{ijt} + \beta_3 dep_{ijt} + \beta_4 age_{ijt} + \beta_5 dis1_{ijt} + \beta_6 dis2_{ijt} + \beta_7 ill1_{ijt} + \beta_8 ill2_{ijt} + \mu_i + \xi_{ijt}$$

$$Y_{ijt} = \beta_1 wait_{ijt} + \beta_2 tiss_{ijt} + \beta_3 dep_{ijt} + \beta_4 age_{ijt} + \beta_5 dis1_{ijt} + \beta_6 dis2_{ijt} + \beta_7 ill1 + \beta_8 ill2_{ijt} + \mu_i + \xi_{ijt}$$

(Model 2)

3.5.3. Econometric model – Mixed Logit.

In order to establish whether there was preference heterogeneity we used Generalized Linear Latent and Mixed Models [GLAMMs] (Rabe-Hesketh, Skrondal et al. 2004). The Mixed Logit model establishes the mean value of coefficients, and then provides a secondary layer of coefficients indicating the standard deviation of each of the regressors. The initial estimating equation (Eberth, Watson et al. 2009) specified has a similar form to Model 2 (although a Mixed Logit model needs to be specified in the econometric package used, which in this case was STATA). The similarity of the estimating model is because Mixed Logit is similar to Conditional Logit except for the fact that parameters are permitted to vary according to prescribed statistical distributions. Therefore preference heterogeneity is incorporated into the model by treating the coefficients as random rather than fixed parameters (Eberth, Watson et al. 2009).

With Mixed Logit an important issue is “choosing the coefficients that are allowed to vary and the distribution they should take” (P1084) (Hole 2008). This means I needed to decide whether all the variables are specified with random coefficients, or some are specified as fixed, and also what distribution to assume for the random parameters.
In relation to the choice of distribution for the random coefficients, the most common assumption when applying Mixed Logit seems to be to assume a normal distribution (Hole 2008) which is what I assumed. However, an alternative might be to assume a log normal distribution (Revelt and Train 1998, Hole 2008) for the error term, and for the random variables. A log normal distribution has the property that “Since the log-normal distribution has positive probabilities only for values greater than zero, specifying a coefficient to be log-normally distributed ensures that it has a positive sign for all individuals. If an attribute is expected to have a negative coefficient (such as waiting time and cost) the attribute is

multiplied by -1 before entering the model and the estimated distribution interpreted as the mirror image of the actual distribution of the coefficient" (Hole 2008). However for our data assuming a log normal distribution would impose unrealistic restrictions upon preferences. This is because a group of DCE respondents do not value the priority criteria as we might expect. It may appear counter-intuitive to prioritize those who have waited a short time rather than those who have waited a long time, or to prioritize those with poor tissue match rather than good tissue match, etc. However, some respondents might do so out of self-interest, because they want to prioritize potential recipients with similar characteristics to themselves. Whilst there may be a case for assuming a log-normal distribution to cater variables for which we would expect all respondents to have the same sign (Revelt and Train 1998), because some patient respondents prioritized potential recipients like themselves whilst others responded altruistically, all our attributes might rationally be valued either positively or negatively. Therefore I assumed a normal distribution for all random variables. Moreover, because all the variables might potentially be subject to preference heterogeneity I specified a model which allowed all the variables to be treated as random. I also assumed a utility model (U_{ijt}) which can be modeled using Y_{ijt} (the dependent variable), which is a function of the variables listed below, and ξ_{ijt} (the error term). We have individuals (i = 1 to n), were 'n' denotes the total number of respondents who choose an alternative (j = 1,..2), in each choice set (t = 1,…16).

$$U_{ijt} = \beta_1 wait_{ijt} + \beta_2 tiss_{ijt} + \beta_3 dep_{ijt} + \beta_4 age_{ijt} + \beta_5 dis1_{ijt} + \beta_6 dis2_{ijt} + \beta_7 ill1_{ijt} + \beta_8 ill2_{ijt} + \xi_{ijt}$$

$$Y_{ijt} = \beta_1 wait_{ijt} + \beta_2 tiss_{ijt} + \beta_3 dep_{ijt} + \beta_4 age_{ijt} + \beta_5 dis1_{ijt} + \beta_6 dis2_{ijt} + \beta_7 ill1_{ijt} + \beta_8 ill2_{ijt} + \xi_{ijt}$$

(Model 3)

3.5.4. Econometric model – Latent Class Model.

After running a Mixed Logit model we ran a LCM (model 4). Like Mixed Logit the LCM technique allows coefficients to vary between respondents (Hole 2008). A good exposition of the statistical methodology underpinning both Mixed Logit and the LCM technique is provided in the discussion paper and published paper produced by Hole (Hole 2007, Hole 2008). So readers with an interest in the underlying statistical methodology are referred to that work. The crucial defining characteristic of the LCM technique (Mentzakis, Ryan et al. 2011) is that the continuous distribution of heterogeneity is approximated by a number of finite 'points of support,' which can be understood as sorting individuals into discrete classes (Greene 2007). However, 'which class contains any particular individual, whether known or not to that individual is unknown to the analyst' (Greene 2007).

In order to determine the optimum number of classes for the Latent Class model we increased the number of classes until the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) suggested we had the optimum number of classes. We continued to increase the number of classes by one class until the AIC and BIC figures stopped falling and began to increase. Both the AIC and BIC reached their lowest level when the model had 4 latent classes, and then began to increase as the number of latent classes increased to 5. Consequently, both the AIC and BIC suggested that the optimum number of latent classes for the Latent Class model was 4. So in the results section we present results for a Latent Class model with 4 latent classes (model 4 in the results section).

This model is very complex and therefore is not readily amenable to being expressed using long hand estimating equations like models 1, 2, 3, and 5. Moreover, it is unlike models 1, 2, 3, and 5, in the sense that the number of classes chosen for analysis is not pre-specified, but

it is determined by increasing the number of classes of an LCM until the Akaike Information Criteria (AIC), and Bayesian Information Criteria (BIC) both reach their lowest level (something that is determined by expanding the number of classes until both the AIC and BIC stop falling and begin to rise).

So given the models complexity I am following the convention of a previous published LCM analysis of DCE data (Grindrod, Marra et al. 2010), in that I only present the details of the estimating equations for more basic models in this methods section. Indeed in their analysis (Grindrod, Marra et al. 2010), the authors only outlined the details of the estimating equation for a more basic multinomial model which they estimated alongside their LCM. However, by setting out the estimating model for their multinomial model first, before conducting their LCM, the authors gave readers an indication of the variables that must be estimated for LCM (although of course the LCM model has to also allow for estimation of coefficients for each of these variables, for each of the latent classes).

Therefore in the same way, I am pointing out here that for our LCM we attempted to obtain coefficient estimates for the value of attribute changes for exactly the same attributes, specified in the equation for the Conditional Logit model (model 2). However the LCM model derives these coefficients separately for each of the 4 latent class groupings. This should be readily apparent from the LCM results reported in section 4.5 of this chapter.

3.5.5. Conditional Logit Model with interaction dummy variables.

Model 5, is a Conditional Logit Model which compares ethnic and non-ethnic minority patient preferences. I decided to use interaction dummy variables to see how preferences varied in the ethnic minority group because I considered that for the reasons discussed in Chapters 5

and 6, there are strong reasons to expect preferences might vary across different ethnic groups of respondents in relation to transplant allocation criteria. Moreover, from a policy making perspective in my view this is probably the most important dimension of preference heterogeneity that requires consideration. In Chapters 5 and 6, I have used Random Effects Probit models. However, because all the models used in this chapter are Logit Models I have used Conditional Logit with dummy variables in this chapter instead (so that for consistency all the models considered in this chapter are Logit models).

The ethnic minority patient category (like that used in Chapter 5) included all patients in an ethnic category except 'White British.' Here, we assume the underlying utility model U_{ijt} is modeled using the dependent variable Y_{ijt} , which is a function of variables listed in the equation below, and ξ_{ijt} (the error term). The term μ_i caters for multiple responses from each respondent using the clustering command in STATA. We have individuals (i = 1 to n), where 'n' denotes the total number of respondents, who choose an alternative (j = 1,..2), in each choice set (t = 1,…16). D_E is a dummy variable, D_E = 1, for ethnic minorities, 0 otherwise.

$$Y_{ijt} = \beta_1 wait_{ijt} + \beta_2 tiss_{ijt} + \beta_3 dep_{ijt} + \beta_4 age_{ijt} + \beta_5 dis1_{ijt} + \beta_6 dis2_{ijt} + \beta_7 ill1_{ijt} + \beta_8 ill2_{ijt}$$
$$+ \beta_9 D_{Eijt} wait_{ijt} + \beta_{10} D_{Eijt} tiss_{ijt} + \beta_{11} D_{Eijt} dep_{ijt} + \beta_{12} D_{Eijt} age_{ijt} + \beta_{13} D_{Eijt} dis1_{ijt} +$$
$$+ \beta_{14} D_{E\,ijt}\, dis1_{ijt} + \beta_{15} D_{E\,ijt}\, ill1_{ijt} + \beta_{16} D_{E\,ijt}\, ill2_{\,ijt} + \mu_i + \xi_{ijt}$$

(Model 5)

Model 5 establishes whether preferences differ between ethnic and non-ethnic minority patients. If any ethnic minority interaction dummies ($\beta_{9....}\beta_{16}$) are significant, they indicate that preferences differ between ethnic minorities and non-ethnic minorities for the associated attribute(s).

3.6. Statistical methods for Marginal Rate of Substitution (MRS).

MRS relates the ratio of changes in other attributes to changes in waiting times (see table 2). In order to make sense of the MRS formula for the LCM analysis, the reader is also referred to the third column in Table 8 (in the LCM model results section), which indicates the coefficient labels used for MRS & Wald tests for each of the coefficients in the estimated model. We used the Delta method (Wooldridge 2002) using command 'nlcom' in STATA, for 95% confidence intervals, to establish the statistical significance or insignificance of MRS.

Wald tests using 'testnl' in STATA for model 4 considered whether MRS differed significantly between the 4 latent classes for a given variable. So for example the Wald test conducted relating to tissue match is $\beta_2 / \beta_1 = \beta_{11} / \beta_{10} = \beta_{20} / \beta_{19} = \beta_{29} / \beta_{28}$.

Wald tests using 'testnl' for model 5 established whether MRS differed significantly between ethnic minority vs. non-ethnic minority patients. Thus to establish (model 5) whether preferences for tissue matching differed between ethnic and non-ethnic minorities, the hypothesis is $\beta_2 / \beta_1 = (\beta_2 + \beta_{10}) / (\beta_1 + \beta_9)$ i.e. was tissue match MRS for non-ethnic and ethnic minorities identical ($p \leq 0.05$ indicates a difference at the 5% level).

4. Results.

4.1. Sample.

Table 3 indicates respondent characteristics. UK Renal Registry data (Byrne, Ford et al. 2008, Byrne, Steenkamp et al. 2008) was used to assess patient sample representativeness. Of the 850 / 863 patients who indicated ethnicity, 762 / 850 patients (89.65%) were white (British), and 24 / 850 (2.82%) were white ethnic minorities, so 92.47% are white. UK

Table 2: Calculating MRS – Valuing attributes compared to a 1 year difference in waiting time .

Models 1, 2, & 3 **Variable**	**MRS**		
wait			
tiss	β_2 / β_1		
dep	β_3 / β_1		
age	β_4 / β_1		
dis1	β_5 / β_1		
dis2	β_6 / β_1		
ill1	β_7 / β_1		
ill2	β_8 / β_1		
Model 4 Variable	**MRS – Latent class 1**	**Model 4 Variable**	**MRS – Latent class 2**
wait		**wait**	
tiss	β_2 / β_1	**tiss**	β_{11} / β_{10}
dep	β_3 / β_1	**dep**	β_{12} / β_{10}
age	β_4 / β_1	**age**	β_{13} / β_{10}
dis1	β_5 / β_1	**dis1**	β_{14} / β_{10}
dis2	β_6 / β_1	**dis2**	β_{15} / β_{10}
ill1	β_7 / β_1	**ill1**	β_{16} / β_{10}
ill2	β_8 / β_1	**ill2**	β_{17} / β_{10}
Model 4 Variable	**MRS – Latent class 3**	**Model 4 Variable**	**MRS – Latent class 4**
wait		**wait**	
tiss	β_{20} / β_{19}	**tiss**	β_{29} / β_{28}
dep	β_{21} / β_{19}	**dep**	β_{30} / β_{28}
age	β_{22} / β_{19}	**age**	β_{31} / β_{28}
dis1	β_{23} / β_{19}	**dis1**	β_{32} / β_{28}
dis2	β_{24} / β_{19}	**dis2**	β_{33} / β_{28}
ill1	β_{25} / β_{19}	**ill1**	β_{34} / β_{28}
ill2	β_{26} / β_{19}	**ill2**	β_{35} / β_{28}
Model 5 Variable	**Non-ethnic minority MRS**	**Model 5 Variable**	**Ethnic minority MRS**
wait		**wait**	
tiss	β_2 / β_1	**tiss**	$(\beta_2 + \beta_{11}) / (\beta_1 + \beta_{10})$
dep	β_3 / β_1	**dep**	$(\beta_3 + \beta_{12}) / (\beta_1 + \beta_{10})$
age	β_4 / β_1	**age**	$(\beta_4 + \beta_{13}) / (\beta_1 + \beta_{10})$
dis1	β_5 / β_1	**dis1**	$(\beta_5 + \beta_{14}) / (\beta_1 + \beta_{10})$
dis2	β_6 / β_1	**dis2**	$(\beta_6 + \beta_{15}) / (\beta_1 + \beta_{10})$
ill1	β_7 / β_1	**ill1**	$(\beta_7 + \beta_{16}) / (\beta_1 + \beta_{10})$
ill2	β_8 / β_1	**ill2**	$(\beta_8 + \beta_{17}) / (\beta_1 + \beta_{10})$

incidence data (Byrne, Ford et al. 2008) suggested 79.7% of renal patients are white, so whites are over-represented in our survey. Overall, 60 / 850 (7.06%) patients indicating ethnicity were non-white, compared with a 20.3% incidence rate (Byrne, Ford et al. 2008)**,** 43 / 60 non-white patients were South Asians (4.98% of those indicating ethnicity) compared to a 10.5% incidence (Byrne, Ford et al. 2008)**.** 480/863 patients (55.62%) were male, 383 / 863 (44.38%) were female. Graphically presented Renal Registry data (Byrne, Ford et al. 2008) reassuringly indicated slightly higher proportions of male than female patients across age

Table 3: Sample characteristics.

	Patients (n = 863)
AGE	
Mean age	54. 75years
GENDER	
Male	480 (55.62%)
Female	383(44.38%)
ETHNICITY	
White (British)	762 (88.19%)
White ethnic minorities	24 (2.78%)
Non-white ethnicity (excluding Asians)	17 (1.97%)
Non-white ethnicity (Asians)	43 (4.98%)
Not indicated	13 (1.51%)
DEPENDENT CHILDREN	
0	718 (83.19%)
1	70 (8.11%)
2	48 (5.56%)
3	12 (41.39%)
> 3	7 (0.81%)
Not indicated	8 (0.93%)
DEPENDENT ADULTS	
0	717 (83.08%)
1	113 (13.09%)
2	16 (1.85%)
> 2	6 (0.70%)
Not indicated	11 (1.27%)

groups. Average sample patient age was 54.75 years (median 56.9 years), and Renal Registry data median age (57.3 years) was virtually identical (Byrne, Ford et al. 2008)**.**

The patient sample comprised: 440 / 863 (50.98%) with successful transplants and 116 / 863 (13.44%) whose transplant failed. Renal Registry prevalence data (Byrne, Steenkamp et al. 2008) suggests 46.9% of patients have successful transplants (close to our figure). There is no incidence / prevalence data for other categories.

4.2. Econometric results - Model 1 (Random Effects Logit).

Table 4: Random Effects Logit (model 1).

Attribute	**Coefficient**	**MRS**	**p-value for MRS**
wait	.1333**		
tiss	.1440**	1.08**	0.000
dep	.1894**	1.42**	0.000
age	--.0101**	-0.08**	0.000
dis1	-.1004*	-0.76*	0.030
dis2	--1.641**	-12.30**	0.000
ill1	..1738**	1.30**	0.004
ill2	.0712	0.53	0.234
Mc Fadden R^2	0.1726	**Proportion of values accurately predicted by the model**	64.20%

***** Indicates significance at the 5% level, but no at the 1% significance level
** Indicates significance at the 1% level.

Results from the Random Effects Logit model 1 (table 4) suggest that patients value a 1% improvement in the 12 month kidney transplant survival rate significantly (MRS=1.08) slightly more than a 1 year additional wait (the denominator for MRS). Patients also value prioritizing those with more dependent adults or children more, and MRS is significant and equals 1.42 for each additional dependent (i.e. those with dependents should wait 1.42 years less for

every dependent they have). Potential transplant recipients who are older are prioritized less (MRS = -0.08). Those with a moderate disease (uncontrolled hypertension or obesity) & Kidney disease affecting life expectancy rather than no disease other than kidney disease are prioritized less than those with no disease affecting life expectancy (apparent because the MRS on dis1 is negative and = -0.76). Likewise those with a severe disease (heart attack, stroke, or diabetes with complications) affecting life expectancy rather than no disease other than kidney disease affecting life expectancy are prioritized far less than those with no disease (apparent because the MRS on dis2 is negative and = -12.30).

Those with moderate disease (mild asthma) rather than no disease affecting quality of life (other than Kidney disease) are prioritized more than those with no disease affecting quality of life (MRS on ill1 is significant and = 1.30), this may be because many patient respondents themselves may have moderate diseases affecting quality of life. The MRS for ill2 is insignificant, suggesting that having severe rather than no disease affecting quality of life should not be a significant determinant of who is prioritized for transplants according to our patient group (a somewhat unexpected result).

4.3. Econometric results - Model 2 (Conditional Logit).

The Conditional Logit model (involving the use of the clustering command to control for multiple responses from each respondent) performs better than Random Effects Logit in terms of measures of goodness of fit (Mc Fadden $R^2$and the proportion of values accurately predicted by the model). Results from model 2, (table 5) again suggest that patients value a 1% improvement in the 12 month kidney transplant survival rate of 1% significantly (MRS = 1.19). Patients also value prioritizing those with more dependent adults or children more, and MRS is significant and equals 1.60 for each additional dependent. Potential transplant

recipients who are older are prioritized less (MRS = -0.09). Those with a moderate disease (uncontrolled hypertension or obesity) & Kidney disease affecting life expectancy rather than no disease other than kidney disease are prioritized less than those with no disease affecting life expectancy (apparent because the MRS on dis1 is negative and = -0.99). Likewise those with a severe disease (heart attack, stroke, or diabetes with complications) affecting life expectancy rather than no disease other than kidney disease affecting life expectancy are prioritized far less than those with no disease (apparent because the MRS on dis2 is negative and = -13.33).The MRS for ill1 and ill2 are both insignificant, suggesting that having moderate or severe diseases affecting quality of life should not be a determinant of who is prioritized for transplants according to our patient group.

Table 5: Conditional Logit (model 2).

Attribute	**Coefficient**	**MRS**	**p-value for MRS**
wait	.0968**		
tiss	.1151**	1.19**	0.000
dep	.1550**	1.60**	0.000
age	-.0090**	-0.09**	0.000
dis1	-.0961*	-0.99*	0.025
dis2	-1.290**	-13.33**	0.000
ill1	-.0206	-0.21	0.719
ill2	-.0050	-0.05	0.931
Mc Fadden R^2	0.2199	**Proportion of values accurately predicted by the model**	71.14%

***** Indicates significance at the 5% level, but no at the 1% significance level
** Indicates significance at the 1% level.

4.4. Econometric results - Model 3 (Mixed Logit).

The Mixed Logit results (model 3, table 6) also suggest that that patients value a 1% improvement in the 12 month kidney transplant survival rate of 1% significantly (MRS=1.24). Patients also value prioritizing those with more dependent adults or children more, and MRS

is significant and equals 1.82 for each additional dependent. Potential transplant recipients who are older are prioritized less (MRS = -0.08). dis1 is insignificant suggesting that those with moderate diseases affecting life expectancy are not prioritized more than those with no diseases affecting life expectancy. However, those with a severe disease (heart attack, stroke, or diabetes with complications) affecting life expectancy rather than no disease other than kidney disease affecting life expectancy are prioritized far less than those with no disease (apparent because the MRS on dis2 is negative and = -13.44). The MRS for ill1 and ill2 are both insignificant, suggesting that having moderate or severe diseases affecting quality of life should not be a determinant of who is prioritized for transplants.

Table 6: Mixed Logit results (model 3)

Attribute	**Coefficient**	**MRS**	**p-value for MRS**
Mean			
wait	.1210**		
tiss	.1506**	1.24**	0.000
dep	.2201**	1.82**	0.000
age	-.0099**	-.08**	0.000
dis1	-.0838	-.07	0.123
dis2	-1.626**	-13.44**	0.000
ill1	-.0400	-.33	0.619
ill2	.0387	.32	0.622
Standard deviation			
wait	-.0568**		
tiss	-.0207		
dep	.2145**		
age	-.0145**		
dis1	.0362		
dis2	1.172**		
ill1	.0957		
ill2	.0159		
Mc Fadden R^2	0.2411	**Proportion of values accurately predicted by the model**	72.78%

***** Indicates significance at the 5% level, but not at the 1% significance level
** Indicates significance at the 1% level.

Most importantly in terms of investigating preference heterogeneity the Mixed Logit results suggest that preference heterogeneity is statistically significant with respect to the variables wait, dep, age, and dis2. Given that wait is the denominator used to calculate MRS, this suggests that preference heterogeneity might impact upon estimated MRS when a Latent Class model is used.

When standard deviations (SD) are significant upon attributes with significant mean values, then if the SD are large in relation to the mean, then this indicates that there is "considerable diversity in the way people value these attributes" (Hall, Fiebig et al. 2006). If you want to cater for diversity you may want to offer a range of provision which varies for those attributes which patients seem to value differently (i.e. with significant standard deviation), particularly if SD is large relative to the mean.

Of importance is the fact that the standard deviations (for those coefficients which are significant) are sometimes quite large in relation to the mean values. For 'wait' the SD = -.0568, which is sizeable compared to the mean .1210. For 'tiss' the SD (.0207) is insignificant, for 'dep' SD (.2145), is a similar size to the mean (.2201); for 'age' SD (-.0145) and mean (-.0099). For dis1 both the mean and SD are insignificant. For 'dis2' the SD (1.172) is sizeable relative to the mean (-1.626). Finally ill1 and ill2 both have an insignificant SD and mean, so these variables do not impact preferences in a significant manner.

It is also possible to use information from the Mixed Logit model to permit calculation of the proportion of the population for whom a treatment characteristic has a positive or negative effect (Hole 2008, Eberth, Watson et al. 2009) if both the mean and standard deviation are significant. This may provide useful information to enable healthcare providers to select the characteristics of service provision in a manner which is utility maximizing for a diverse

patient population. So using the significant results from the basic Mixed Logit Model (table 5), For ‘wait’ 9.7% valued it negatively and 90.3% valued it positively; for ‘dep’ 16.8% valued it negatively and 83.2% valued it positively; for ‘age’ 68.9% valued it negatively and 31.1% valued it positively; for ‘dis2’ 86.3% valued it negatively and 13.7% valued it positively. However, such figures may simply be an *artefact* of the assumption of normality (Hole 2008), so if the assumption of normally distributed random parameters does not hold neither will such figures. To be confident these results might be valid, we would want to ensure a scale heterogeneity model was not a better fit for the data (Fiebig, Keane et al. 2010), and also that the assumption of normality for the random parameters was valid.

Often if a cost variable is included in analysis it is treated as fixed, because “willingness to pay for each attribute (which is the ratio of the attribute’s coefficient to the price coefficient) is thereby distributed in the same way as the attribute’s coefficient” (Revelt and Train 1998). This “ensures that the coefficient has the right sign: a normally distributed cost coefficient implies that some individuals may prefer an appointment with higher cost which is counter-intuitive” (Hole 2008). However for this DCE we did not have a cost attribute. Rather instead we had a variable relating to waiting time for a transplant, used as the denominator for MRS (Marginal Rate of Substitution).

When significant preference heterogeneity in terms of cost cannot be ruled out, there is evidence to suggest that models where the data is allowed to vary may fit the data better (Meijer and Rouwendal 2006, Hole 2008). Although this DCE did not have a cost attribute we had to decide whether to treat the value of the ‘wait’ attribute (used to derive MRS) as fixed or to allow it to vary. Since significant preference heterogeneity in relation to the ‘wait’ attribute could not be ruled out, and some respondents would value differences in ‘wait’ positively and others would value it negatively, we allowed the variable ‘wait’ to vary like all the others.

4.5. Econometric results - Model 4 (Latent Class Model).

Table 7: Latent Class Model results (AIC and BIC) according to the number of classes in the model.

	Class 1	Class 2	Class 3	Class 4	Class 5
AIC	8415.54	7889.75↓	7714.96↓	7482.22↓	7545.94↑
BIC	8476.75	8019.81↓	7913.88↓	7750.00↓	7882.57↑

We kept running Latent Class models with increasing numbers of latent classes until both the AIC and BIC figures ceased to improve. Both figures continued to improve (i.e. declined) until the model had 4 classes (table 7). They then began to deteriorate when a model was applied with 5 latent classes. We therefore used a model with 4 latent classes, and some of the results are presented in table 8a. Important results presented in table 8a are the fact that the probability of belonging to latent class 1 is 0.3017; probability of belonging to latent class 2 is 0.1983; probability of belonging to latent class 3 is 0.1768; and the probability of belonging to latent class 4 is 0.3233. This means that taking the sample as a whole there are different probabilies that a respondent will belong to each of the 4 latent class groups (each of which are associated with different coefficients for the attributes).

In table 8a the results of the econometric analysis which involves 4 latent classes are presented. In order that it can be readily seen which results relate to which latent class, the latent class a particular coefficient relates to is indicated by the terms '_lc1' or '_lc2' or '_lc3' or '_lc4' after the variable names. This is in order to indicate whether coefficients relate to latent classes 1, 2, 3, or 4 respectively.

MRS figures suggest (table 8a) that MRS with respect to tissue match (the value of a 1% change in the likelihood of a kidney transplant succeeding for 12 months or more compared

to someone waiting one year longer) is always significant and positive, and varies from a 1.28 low, to a 1.73 high across the 4 latent classes suggesting that patients value prioritizing those with a better tissue match between recipient and donor. However, Wald test results (table 9) do not support the hypothesis that MRS for tissue match varies across latent classes (p=0.4069). The MRS relating to prioritizing those with an extra dependent adult or child varies from a low of 0.20 to a high of 3.70 across the 4 latent classes, so MRS is very sensitive to class membership for prioritizing those with more dependents, implying that there is quite a lot of preference heterogeneity with respect to valuing having dependents. This is confirmed by the Wald test (p=0.000) which suggests that MRS for 'dep' varies across latent classes. MRS relating to the age of recipients varies. Intuitively we would expect people to prioritize younger rather than older recipients (because younger recipients generally have more potential to benefit from a transplant) so we expect the sign on MRS with respect to being a year older to be negative. Interestingly however the coefficient changes sign across latent classes. It is positive and significant for classes 1 and 2 (0.13 and 0.11 respectively) and becomes negative and significant in classes 3 and 4 (-0.54 and -0.042 respectively). This preference heterogeneity is confirmed by the Wald test for MRS for 'age' which confirms that MRS differs across latent classes (p=0.000). The MRS for 'dis1' is not significant in latent classes 1 and 4. In latent class 2, MRS is positive and significant at 5.32 and in latent class 3 it is also positive and significant but higher at 14.52. Wald test results suggest that MRS for 'dis1' varies across latent classes (p=0.000). MRS for dis2 is always significant. In latent class 2 it is positive at 3.23, but in classes 1, 3, and 4 it is negative, it is -24.59 in latent class 1, -11.10 in latent class 3, and -5.36 in latent class 4. Wald test results suggests that MRS for 'dis2' varies across latent classes (p=0.000).

Table 8a: Latent Class Model (model 4) results for 4 Latent Classes.

Attribute	Coefficient	Coefficient label for MRS in table 2, and the Wald test in table 9	MRS	p-value
wait_lc1	.1986**	β_1		
tiss_lc1	.2541**	β_2	1.28**	0.000
dep_lc1	.1465**	β_3	0.73*	0.012
age_lc1	.0254	β_4	0.13**	0.008
dis1_lc1	.1421	β_5	0.72	0.541
dis2_lc1	-4.883**	β_6	-24.59**	0.000
ill1_lc1	2.219*	β_7	11.17**	0.000
ill2_lc1	1.157	β_8	5.83*	0.014
P2_1	-.0691			
wait_lc2	4.198**	β_{10}		
tiss_lc2	6.673**	β_{11}	1.59**	0.000
dep_lc2	0.853**	β_{12}	0.20**	0.000
age_lc2	0.441**	β_{13}	0.11**	0.000
dis1_lc2	22.34**	β_{14}	5.32**	0.000
dis2_lc2	13.56**	β_{15}	3.23**	0.000
ill1_lc2	22.47**	β_{16}	5.35**	0.000
ill2_lc2	41.91**	β_{17}	9.98**	0.000
P2_2	-.489**			
wait_lc3	0.252**	β_{19}		
tiss_lc3	0.435**	β_{20}	1.73**	0.000
dep_lc3	0.648**	β_{21}	2.57**	0.000
age_lc3	-0.136**	β_{22}	-0.54**	0.000
dis1_lc3	3.659**	β_{23}	14.52**	0.000
dis2_lc3	-2.798**	β_{24}	-11.10**	0.000
ill1_lc3	-0.253	β_{25}	-1.00	0.535
ill2_lc3	1.50**	β_{26}	6.11**	0.002
P2_3	-.0604**			
wait_lc4	.064**	β_{28}		
tiss_lc4	.0838**	β_{29}	1.30**	0.000
dep_lc4	0.2388**	β_{30}	3.70**	0.000
age_lc4	-0.003	β_{31}	-.042	0.329
dis1_lc4	-0.164*	β_{32}	-2.55	0.069
dis2_lc4	-0.346*	β_{33}	-5.36**	0.002
ill1_lc4	-0.250*	β_{34}	-3.87	0.051
ill2_lc4	-0.152	β_{35}	-2.36	0.205
Probability class 1:	0.3017		**Probability class 2:**	0.1983
Probability class 3:	0.1768		**Probability class 4:**	0.3233
Mc Fadden R^2	0.3115		**Proportion of values accurately predicted by the model**	83.48 %

***** Indicates significance at the 5% level, but not at the 1% significance level

** Indicates significance at the 1% level.

Table 8b: Latent Class Model (model 4) – Class assignment model information.

	Class 1	Class 2	Class 3	Class 4
Probability of belonging to the class	0.3017	0.1983	0.1768	0.3233
Variable relating to patient respondent characteristics	**Class 1**	**Class 2 coefficients**	**Class 3 coefficients**	**Class 4 coefficients**
constant	Base class	.955**	1.514**	.279**
ethnic_min	Base class	.099	.487**	.762**
gender _male	Base class	-.333**	.035	-.099**
age	Base class	-.021**	**-**.038**	-.005**
dep_children	Base class	-.243**	-.225**	-.099
dep_adults	Base class	.201**	.066	.251**

* Indicates significance at the 5% level, but not at the 1% significance level.
** Indicates significance at the 1% level.

Table 9: Latent Class Model (model 4) Wald test results for MRS.

MRS	Wald test hypothesis	p-value of Wald test
tiss	$\beta_2 / \beta_1 = \beta_{11} / \beta_{10} = \beta_{20} / \beta_{19} = \beta_{29} / \beta_{28}$	0.4069
dep	$\beta_3 / \beta_1 = \beta_{12} / \beta_{10} = \beta_{21} / \beta_{19} = \beta_{30} / \beta_{28}$	0.0000
age	$\beta_4 / \beta_1 = \beta_{13} / \beta_{10} = \beta_{22} / \beta_{19} = \beta_{31} / \beta_{28}$	0.0000
dis1	$\beta_5 / \beta_1 = \beta_{14} / \beta_{10} = \beta_{23} / \beta_{19} = \beta_{32} / \beta_{28}$	0.0000
dis2	$\beta_6 / \beta_1 = \beta_{15} / \beta_{10} = \beta_{24} / \beta_{19} = \beta_{33} / \beta_{28}$	0.0000
ill1	$\beta_7 / \beta_1 = \beta_{16} / \beta_{10} = \beta_{25} / \beta_{19} = \beta_{34} / \beta_{28}$	0.0000
ill2	$\beta_8 / \beta_1 = \beta_{17} / \beta_{10} = \beta_{26} / \beta_{19} = \beta_{35} / \beta_{28}$	0.0000

The MRS for 'ill1' is significant in latent classes 1 and 2 but insignificant in classes 3 and 4. In latent class 1 it is 11.17, and in latent class 2 it is 5.35. Wald test results suggest that MRS for 'ill1' varies significantly across latent classes (p=0.000). MRS for 'ill2' is significant in classes 1, 2, and 3. It is 5.83 in class 1, 9.98 in class 2, and 6.11 in class 3, but insignificant in class 4. Wald test results also suggest that MRS for 'ill2' varies significantly across latent classes (p=0.000). Therefore the general picture emerging from the 4 class Latent Class model is that MRS differs in a statistically significant manner across latent classes for every variable except

'tiss' (tissue match). This suggests that adopting econometric modeling which allows for preference heterogeneity is appropriate.

Information in table 8b is particularly helpful as it indicates whether the patient characteristics in question influence the likelihood of being in a particular latent class. With respect to the class assignment model variable relating to ethnic minorities (ethnic_min), the dummy variable (=1 for ethnic minorities and 0 otherwise) is not significantly different in class 2 relative to class 1. However, it is significantly different from class 1 in class 3 (coefficient = .487) and also for class 4 (coefficient = .762). The coefficients in classes 3 and 4 are both positive which implies that individuals in classes 3 and 4 are more likely to be ethnic minorities. It is important to note that the valuation of the tissue match coefficient (.0838) is lower in class 4 than for all the other latent classes (class 1 = .2541, class 2 = 6.673, class 3 = 0.435). Moreover class 4 is the class with the highest proportion of ethnic minorities in it. This suggests that the presence of ethnic minorities in this latent class group may serve to lower the average respondents valuation of the tissue match attribute within that latent class group. This fits with the hypothesis that ethnic minorities may disadvantage themselves if they value a close tissue match between donor and recipient positively. Somewhat counter-intuitively however, it would appear that the valuation of having a close tissue match between donor and recipient (tiss) is higher in class 3 (coefficient = 0.435) than in class 1 (.2541), when class 3 members are statistically significantly more likely to be ethnic minorities than class 1 members. This means that the LCM model fails to demonstrate a clear cut relationship between the ethnic minority status of respondents and them valuing a close tissue match between donor and recipient by less than non-ethnic minorities. In contrast a more clear-cut relationship between these variables is illustrated in model 5 below.

In relation to the dummy variable (gender _male) relating to gender (=1 for male respondents and 0 otherwise), the dummy variable is not significantly different from class 1 in class 3. However, it is significantly different and negative in class 2 (coefficient = -.333) and in class 4 (coefficient= -.099), which implies that individuals in classes 2 and 4 are less likely to be male.

In relation to the variable relating to respondents actual age (age), the coefficients for classes 2 (coefficient = -.021), 3 (=-.038), and 4 (= -.005) all differ significantly, and are all negative relative to class 1. This implies just like the published LCM analysis (Mentzakis, Ryan et al. 2011) which also had negative significant variables relating to age, that individuals in these classes (classes 3 and 4) are more likely to be younger.

In relation to the dummy variables relating to whether respondents have dependent children (dep_children), this dummy variable is not significantly different from class 1 in class 4. However, in both class 2 (coefficient = -.243) and class 3 (coefficient = -.225), it is significantly different and negative relative to class 1.This implies that individuals in classes 2 and 3 are less likely to have dependent children.

In relation to the dummy variables relating to whether respondents have dependent adults (dep_adults), this dummy variable is not significantly different from class 1 in class 3. However, it is significantly different and positive in class 2 (coefficient =.201) and class 4 (coefficient = .251). This implies that individuals in classes 2 and 4 are more likely to have adult dependents to look after.

The relevance of latent classes to the policy maker are that the latent classes allow policy makers to be aware that within an overall pooled DCE sample there may exist different

classes of respondents who value attributes differently. If a policy maker wants to tailor healthcare provision to cater for diversity he or she could attempt to ensure that healthcare provision is tailored to the needs of respondents from each of the latent classes identified in a DCE. Also, because information on the percentage of the whole sample falling into each of the latent classes is generated, this allows the policy maker who wants to customize healthcare provision to meet the needs of different latent class groups, to know what proportion of people would prefer provision characterized by different provision tailored to the preferences of different latent class groups. For example an LCM analysis (Grindrod, Marra et al. 2010) found that for the sample of 87 students they had 2 latent classes with 50% in each group, the first valued the service type attribute over and above everything else, and preferred to work in hospital pharmacies or research internships, whilst the other class of students valued personal income and job satisfaction above all the other attributes. Those arranging work placements for students could therefore use this information to tailor provision for the 2 different classes of students accordingly. In a similar way other latent class DCE analyses (Hole 2008, Mentzakis, Ryan et al. 2011, Mentzakis, Stefanowska et al. 2011, Waschbusch, Cunningham et al. 2011, Naik-Panvelkar, Armour et al. 2012, Carroll, Al-Janabi et al. 2013, de Bekker-Grob, Rose et al. 2013, Whitty, Stewart et al. 2013) define different numbers of latent classes, and the analyses show that classes vary in the valuation of different attributes.

The relevance of having 4 latent classes for this DCE is that it can be seen that in a relatively large sample of respondents (n = 863), 4 different latent classes of respondents can be defined who value the attributes relating to priority criteria for renal transplantation differently. The fact that we have 4 latent classes is not surprising because we have a relatively large sample of DCE respondents. Each of the latent classes is defined by different coefficients for the attributes (table 8a). Our LCM analysis (which incorporates class assignment prediction

modelling), also allows us to establish whether particular latent classes are more or less likely to be categorized by certain respondent characteristics (table 8b). The findings suggest that compared to class 1 (who make up about 30% of the sample, probability = 0.3017), class 2 members (who make up about 20% of the sample, probability = 0.1983) are less likely to be male, less likely to be older (and hence more likely to be younger), less likely to have dependent children, but more likely to have dependent adults to care for. Also compared to class 1, class 3 members (who make up about 18% of the sample, probability = 0.1768) are more likely to be ethnic minorities, less likely to be older (and hence more likely to be younger), and less likely to have dependent children. Finally, compared to class 1, class 4 members (who make up about 32% of the sample, probability = 0.3233) are more likely to be ethnic minorities, less likely to be male, less likely to be older (and hence more likely to be younger), and more likely to have dependent adults to care for.

Crucially however because the objective of renal transplant policy is to establish a set of allocation criteria for the UK overall, it is unlikely that a policy could be derived for different latent classes of patient preference simultaneously. So in this policy context, the main use of the latent class information is to establish those attributes for which patient valuations may change sign, or for which the valuation of attributes varies a lot in magnitude. It could be that criteria that appear to be valued very differently across the latent classes may be weaker candidates to use as allocation criteria, compared to those attributes which never change sign, or do not vary as much in magnitude.

4.6. Econometric results - Model 5 (Conditional Logit with interaction dummy variables for ethnic minorities).

Results for model 5 (table 10) which involves the application of conditional logit using the clustering command, suggest that MRS differs in a statistically significant manner between ethnic minorities and non-ethnic minorities with respect to tiss, dep, age, dis1 and dis2. Non-ethnic minority respondents value prioritizing recipients with a closer tissue match positively, but ethnic minorities do not (MRS is insignificant). Non-ethnic minorities would prioritize recipients with dependents (dep is positive and significant), but amongst ethnic minorities MRS for dep is insignificant. Non-ethnic minorities would prioritize older recipients less (MRS for age is negative and significant), but amongst ethnic minorities MRS for age is insignificant. Non-ethnic minorities have a negative MRS for dis1 which is -1.34, but ethnic minorities have a positive MRS of 2.61. However, both figures for MRS are insignificant, but the Wald test does suggest that MRS differs significantly between the 2 groups. For dis2, MRS is negative and significant amongst non-ethnic minorities (-14.05) amongst ethnic minorities though it is positive and significant (6.74). This suggests that non-ethnic minorities would not prioritize those with severe disease affecting life expectancy over those with no disease affecting life expectancy, whereas ethnic minorities would prioritize those with severe diseases affecting life expectancy over those with no disease (perhaps because ethnic minorities are more likely to suffer from severe diseases affecting life expectancy). Amongst both non-ethnic minorities and ethnic minorities ill1 and ill2 is insignificant (also Wald tests do not suggest that MRS differs between non-ethnic minorities and ethnic minorities with respect to ill1 or ill2).

Table: 10: Conditional Logit with interaction dummy variables (model 5) for ethnic minority groups (88 out of 863 are ethnic minorities).

Variable	Coefficient for non-ethnic minorities	MRS for non-ethnic minorities.	Coefficient for dummy variables for ethnic minorities	MRS for ethnic minority patients	Wald test p-values
wait	.0099**	1	-.0174	1	
tiss	.1242**	1.25** (0.97 / 1.53)	-.0609*	0.51 (-1.08 / 0.68)	p=0.0000
dep	.1584**	1.59** (1.29 / 1.90)	-.0114	0.026 (-0.70/ 0.75)	p=0.0004
age	-.0096**	-0.10** (-.13 / -.07)	.0040	0.13 (-.08 / 0.10)	p=0.0443
dis1	-.1333**	-1.34** (-2.25 / -.43)	-.2748*	2.61 (-0.03 / 5.25)	p=0.0115
dis2	-1.397**	-14.05** (-12.17/ -15.92)	.8108**	6.74** (3.35 / 10.14)	p=0.0000
ill1	-.0324	-0.33 (-1.56 / 0.91)	.0650	0.60 (-2.47 / 3.68)	p=0.6235
ill2	-.0440	-0.44 (-1.71 / 0.83)	.2359	2.31 (-0.815 / 5.44)	p=0.1549
Mc Fadden R^2	0.2235		**Proportion of values accurately predicted by the model**	**73.69%**	

***** Indicates significance at the 5% level, but no at the 1% significance level
** Indicates significance at the 1% level.

5. Discussion.

Baseline findings which do not consider whether preferences are heterogeneous or not (models 1 and 2) are broadly supportive of 2006 revisions to UK kidney transplant policy which prioritized long waiters and young adults. These analyses indicate that this shift in policy was justified, but suggest that other criteria (i.e. prioritizing those with dependents) ought to be considered when UK transplant policy is re-appraised.

Models 3 and 4 (Mixed Logit and Latent Class) are general models which can be used to assess the extent to which preferences are heterogeneous. Both analyses highlight the fact that preferences for some variables are heterogeneous. The Mixed Logit results indicate that preference are statistically significantly different (heterogeneous) with respect to 4 / 8 of the variables (wait, dep, age, and dis2) indicated by a statistically significant measure of standard deviation.

For the Latent Class Model a slightly different measure of heterogeneity was deployed. Here we considered it logical to normalize our valuation of all of the other variables in terms of waiting time (wait) as in Chapters 5 and 6. So we used 'wait' which relates to 1 year of waiting time for the denominator to derive MRS. We then used Wald tests to test for equality of MRS for each variable across the 4 latent classes.

Both the Mixed Logit and the LCM results in table 8a and 9 failed to provide strong support for the view that preferences for 'tiss' (prioritizing those with a close tissue match between recipient and donor) were heterogeneous. For the Mixed logit model, the coefficient relating to preference heterogeneity, i.e. SD on 'tiss' (.0207) was insignificant, Also the Wald test

results (table 9) did not provide any evidence to support the view that the valuation of the 'tiss' attribute varied in a statistically significant manner across the 4 latent classes.

The evidence from the LCM class assignment model (table 8b) fails to provide clear-cut evidence that ethnic minorities valued the closeness of tissue match variable (tiss) less than non-ethnic minorities. Class 4 latent class members have the largest positive coefficient relating to ethnic minority status (.762), and they are statistically significantly more likely to be ethnic minorities than class 1 members, they also have a lower valuation of the 'tiss' variable than any other group (.0838). This fits with our hypothesis that ethnic minorities value the 'tiss' variable less than non-ethnic minorities. However, class 3 members are also significantly more likely to be ethnic minorities (coefficient = .487) than class 1 members, but they value the closeness of tissue match variable more highly (coefficient =.435) in class 3, than those in class 1 (coefficient =.2541).

The Wald test results (table 9) for the Latent Class Model (model 4) suggested that preferences differed across latent classes for all the other variables (i.e. 6 / 7 of the measures of MRS). This is in contrast to the results from the conditional logit model with dummy variables which suggested that preferences for 'tiss' varied according to ethnicity. That said both the Mixed Logit and LCM uncovered evidence of unobserved preference heterogeneity, which the conditional logit with dummy variables model (model 5) could not uncover.

If you take the findings from the Mixed logit and LCM at face value, they might be considered to imply that the Latent Class results indicate more preference heterogeneity than Mixed Logit results (for which only 4 / 8 variables had a statistically significant standard deviation).

This finding should not be unexpected. This is because our Mixed Logit results indicated that there is preference heterogeneity with respect to the variable ‘wait’. Since the variable ‘wait’ is the denominator that was used to derive MRS for the Latent Class Model, it follows that we always use a denominator to derive MRS which is subject to preference heterogeneity. Moreover, since our Wald tests for MRS look at variation across 2 variables, whereas the Mixed Logit standard deviation measure of heterogeneity assesses heterogeneity only with respect to one variable, it is hardly surprising that our Wald test measures for the 4 class Latent Class Model seem to imply more evidence of heterogeneity than Mixed Logit results.

Mixed Logit has the advantage that it is simpler to undertake that LCM, and the results generated are shorter and hence easier to interpret. However, the main problem with the method is that for the results to be valid your distributional assumptions have to be valid. Although I consider that there are strong grounds for considering that the use of a log distribution is less appropriate than assuming that preference heterogeneity is normally distributed (our assumption) for this DCE application, I cannot rule out the possibility that in reality preference heterogeneity might exhibit a skewed rather than normal distribution. If this was the case then the results of the Mixed Logit (applied here) which assume a normal distribution for preference heterogeneity will not be valid. Moreover, unless I run a scale heterogeneity model on this data (Fiebig, Keane et al. 2010) for comparison, I cannot be sure that some of the heterogeneity that may be attributed to preference heterogeneity might in reality be attributable to scale heterogeneity. There is some support for these concerns, in a comparison of results from a Juster scale with a Mixed Logit (MXL) for a health application, the analysis concluded that findings lent “support to the findings of Fiebig and colleagues and suggests that MXL may not be the optimal model for this dataset.”(Whitty, Rundle-Thiele et al. 2012).

On balance comparing Mixed Logit and LCM, I consider that the results from the LCM model may be more accurate (because such analysis does not require the imposition of distributional assumptions relating to heterogeneity). For the Latent Class Model a slightly different measure of heterogeneity was deployed. Here we considered it logical to normalize our valuation of all of the other variables in terms of waiting time (wait) like all the other models in this paper. So we used 'wait' which relates to 1 year of waiting time for the denominator to derive MRS. However, in contrast to the Mixed Logit model we then used Wald tests to test for equality of MRS for each variable across the 4 latent classes. Neither the Mixed Logit or LCM results provide much support for the view that preferences for 'tiss' (prioritizing those with a close tissue match between recipient and donor) were heterogeneous in a clear cut manner, a finding which conflicts with the findings of the conditional logit with dummy variables model (model 5).

An advantage of the LCM is that distributional assumptions do not have to be made, because the discrete distributions in the latent class model can be interpreted as non-parametric estimates of the continuous distributions in the mixed logit model (Hole 2008). Whilst a difficulty remains in terms of determining the optimal number of latent classes with latent class analysis, criteria such as Bayesian and Akaike information criteria can be used to help inform the optimal number of latent classes.

One reason why Mixed Logit rather than LCM has been used more for DCEs, is that it is computationally easier to undertake. However, as already pointed out information obtained from LCM models (Grindrod, Marra et al. 2010) may have the advantage that it can be used to customize health care provision to meet the needs of different latent classes of service users. Although the Mixed Logit approach is less powerful in this respect, if valid distributional

assumptions are applied it may help you identify the fact that certain attributes are subject to heterogeneity using a less complicated model than LCM.

The policy implications of the these findings are that Models 1 and Model 2 which did not cater for preference heterogeneity at all inevitably failed to identify preference heterogeneity exposed by Mixed Logit, LCM, and the conditional logit model with dummy variables. For the methodological reasons I have just outlined I am unsure about how reliable the findings from the Mixed Logit model are. This means that the policy making value of this analysis, and also the information on the percentage of the sample (for variables with significant standard deviations) with negative as opposed to positive coefficients may simply be an artifact of inappropriate distributional assumptions. Therefore if the distributional assumptions are incorrect then the policy making value of the findings is questionable. Without comparing Mixed Logit preference heterogeneity model with a scale heterogeneity model (to see which is a better fit) I may be attributing some of the heterogeneity to preference heterogeneity using Mixed Logit when this is incorrect (Fiebig, Keane et al. 2010).

In contrast the LCM model has the advantage that it does not require us to make distributional assumptions about preference heterogeneity, just the number of classes included in the model (a process that can be guided using Bayesian and Akaike information criteria). So its policy making utility is not particularly questionable on the grounds that the results may be invalid because of the potential imposition of inappropriate assumptions about the distribution of preference heterogeneity.

Like the Mixed Logit Model, the LCM did not identify any clear cut pattern of preference heterogeneity with respect to the ‘tiss’ (closeness of tissue match) variable (which could be related to the ethnic minority status of respondents). In contrast the conditional logit model

with interaction dummy variables model did allow us to reach the very policy relevant conclusion that the 'tiss' variable was valued positively and significantly amongst non-ethnic minorities (coefficient on 'tiss' was .1254, with a significant MRS of 1.25) however amongst ethnic minorities both the coefficient on the variable and MRS were insignificant, and the Wald test indicated a significant difference in MRS for this variable between the groups. Of course however, the conditional logit with dummy variables model is not equipped to identify unobserved preference heterogeneity, whereas the Mixed Logit and LCM models did provide information about this.

6. Conclusions.

The main finding to highlight, emerging from this analysis, is the fact that whilst both the Mixed Logit and LCM results provide no evidence of preference heterogeneity with respect to the tissue match attribute (tiss), this is at odds with the findings of the Conditional Logit model with dummy variables (model 5). Interestingly, model 5 has dummy variables for ethnic minority patient respondents. Many ethnic minorities would be disadvantaging themselves as an ethnic group if they favoured prioritizing recipients with a close tissue match between donor and recipient. This is because low levels of organ donation amongst some ethnic minority groups, mean that these populations are far less likely to be able to get a closely tissue matched transplant. Therefore, the fact that model 5 results show that non-ethnic minorities would prioritize recipients with a close donor tissue match (MRS = 1.25 for 'tiss' and highly significant) whereas MRS for 'tiss' is insignificant amongst ethnic minorities (a difference also highlighted by the Wald test result [p=0.000]) makes intuitive sense. It is probably the most noteworthy finding in relation to preference heterogeneity, both in the analysis in model 5 reported here, and also in the earlier analysis reported in Chapters 5 and 6.

However, had we not employed an interaction dummy variable model such as model 5 to establish whether preferences differ between non-ethnic minorities and ethnic minorities, we may have reached the misleading conclusion (based only on Mixed Logit and LCM results without interaction dummy variables) that there is no clear-cut evidence of preference heterogeneity with respect to the tissue match attribute. This finding suggests that we cannot just rely upon blanket application of Mixed Logit or LCM analysis in order to pinpoint variations in preferences. Instead, if we anticipate that preferences might vary between defined respondent groups, we ought to use dummy variable models to test for this perhaps within a Mixed Logit model, or an LCM model . The model 5 results in this paper, and those reported in our earlier analysis (see Chapters 5 and 6), suggest that both time spent waiting and the quality of tissue match between donor and recipient are of importance to non-ethnic minority patients, but that amongst ethnic minority patients closeness of tissue match is not a significant determinant of patient preferences. These findings highlight significant issues about kidney transplant allocation to those from ethnic minority groups.

In total we obtained 908 patient responses, and 863 of these proved to be complete enough to be amenable to data analysis using all 5 econometric models. Both Mixed Logit and LCM analysis highlighted the fact that preferences appear to be heterogeneous across many variables, and identified a lot of unobserved preference heterogeneity. This was not a surprising result given that we had such a large sample for analysis. However, the Mixed Logit and LCM applied, proved insufficiently sensitive to highlight any preference heterogeneity with respect to the variable ‘tiss’. Only when a model using interaction dummy variables for ethnic minority groups was used, did a clear cut pattern of preference heterogeneity for this attribute between ethnic groups become apparent. We would therefore

suggest that deploying interaction dummy variable models can be particularly useful in that they provide an extra layer of information.

The main purpose of the analysis presented in this chapter was to demonstrate that the blanket application of either Mixed Logit or LCM techniques to DCE response data is no substitute for giving adequate forethought prior to data analysis as to whether or not sub-groups of respondents might be expected to have different preferences. If differences are anticipated then, by using interaction dummy variables, or establishing how preferences might vary for sub-groups of respondents using LCM, any differences in preferences for defined respondent sub-groups can potentially be established. Indeed, in terms of 'policy making utility', arguably the findings of the Conditional Logit model with interaction dummy variables conferred more information of policy significance about preference heterogeneity than the Mixed Logit without dummy variables or our LCM model .

7. References.

Amaya-Amaya, M. and K. Gerard (2008). "Discrete Choice Experiments in a Nutshell." Chapter 1, Using discrete choice experiments to value health and care: p13-46.

Blaauw, D., E. Erasmus, N. Pagaiya, V. Tangcharoensathein, K. Mullei, S. Mudhune, C. Goodman, M. English and M. Lagarde (2010). "Policy interventions that attract nurses to rural areas: a multicountry discrete choice experiment." Bull World Health Organ **88**(5): 350-356.

Byrne, C., D. Ford, J. Glig, D. Ansell and J. Feehally (2008). "ESRD incident rates in 2008: national and centre-specific analyses." **Chapter 3, UK Renal Registry Report [http://www.renalreg.org]**.

Byrne, C., R. Steenkamp, C. Castledine, D. Ansell and J. Feehally (2008). "ESRD prevalent rates in 2008 national and centre-specific analyses." **Chapter 4, UK Renal Registry report [http://www.renalreg.org]**.

Carroll, F. E., H. Al-Janabi, T. Flynn and A. A. Montgomery (2013). "Women and their partners' preferences for Down's syndrome screening tests: a discrete choice experiment." Prenat Diagn **33**(5): 449-456.

Daly, A., S. Hess and K. Train (2012). "Assuring finite moments for willingness to pay in random coefficient models." Transportation **39**(1): 19-31.

de Bekker-Grob, E. W., M. L. Essink-Bot, W. J. Meerding, H. A. Pols, B. W. Koes and E. W. Steyerberg (2008). "Patients' preferences for osteoporosis drug treatment: a discrete choice experiment." Osteoporos Int **19**(7): 1029-1037.

de Bekker-Grob, E. W., R. Hofman, B. Donkers, M. van Ballegooijen, T. J. Helmerhorst, H. Raat and I. J. Korfage (2010). "Girls' preferences for HPV vaccination: a discrete choice experiment." Vaccine **28**(41): 6692-6697.

de Bekker-Grob, E. W., J. M. Rose, B. Donkers, M. L. Essink-Bot, C. H. Bangma and E. W. Steyerberg (2013). "Men's preferences for prostate cancer screening: a discrete choice experiment." Br J Cancer **108**(3): 533-541.

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

Eberth, B., V. Watson, M. Ryan, J. Hughes and G. Barnett (2009). "Does one size fit all? Investigating heterogeneity in men's preferences for benign prostatic hyperplasia treatment using mixed logit analysis." Med Decis Making **29**(6): 707-715.

Fiebig, D. G., M. P. Keane, J. Louviere and N. Wasi (2010). "The Generalized Multinomial Logit Model: Account for Scale and Coefficient Heterogeneity." Marketing Science **29**(3): 393-421.

Fiebig, D. G., M. P. Keane, J. Louviere and N. Wasi (2010). "The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity." Marketing Science **29**(3): 393-421.

Goto, R., S. Nishimura and T. Ida (2007). "Discrete choice experiment of smoking cessation behaviour in Japan." Tobacco Control **16**(5): 336-343.

Goto, R., Y. Takahashi and T. Ida (2011). "Changes in Smokers' Attitudes Toward Intended Cessation Attempts in Japan." Value in Health **14**(5): 785-791.

Greene, W. H. (2007). "Limdep v9.0 - Econometric Modelling Guide." Econometric Software, Inc **Vol. 1.**(New York).

Grindrod, K. A., C. A. Marra, L. Colley, R. T. Tsuyuki and L. D. Lynd (2010). "Pharmacists' preferences for providing patient-centered services: a discrete choice experiment to guide health policy." Ann Pharmacother **44**(10): 1554-1564.

Guimaraes, C., C. A. Marra, L. Colley, S. Gill, S. H. Simpson, G. S. Meneilly, R. H. Queiroz and L. D. Lynd (2009). "A valuation of patients' willingness-to-pay for insulin delivery in diabetes." Int J Technol Assess Health Care **25**(3): 359-366.

Guimaraes, C., C. A. Marra, S. Gill, S. Simpson, G. Meneilly, R. H. Queiroz and L. D. Lynd (2010). "A discrete choice experiment evaluation of patients' preferences for different risk, benefit, and delivery attributes of insulin therapy for diabetes management." Patient Prefer Adherence **4**: 433-440.

Hall, J., D. G. Fiebig, M. T. King, I. Hossain and J. J. Louviere (2006). "What influences participation in genetic carrier testing? Results from a discrete choice experiment." J Health Econ **25**(3): 520-537.

Hauber, A. B., A. F. Mohamed, F. R. Johnson and H. Falvey (2009). "Treatment preferences and medication adherence of people with Type 2 diabetes using oral glucose-lowering agents." Diabet Med **26**(4): 416-424.

Hoffman, S. D. and G. J. Duncan (1988). "Multinomial and Conditional Logit Discrete-Choice Models in Demography." Demography **25**(3): 415-427.

Hole, A. R. (2007). "Fitting mixed logit models using maximum simulated likelihood." The Stata Journal **7**(3): 388-401.

Hole, A. R. (2007). "Modelling heterogeneity in patients' preferences for the attributes of a general practitioner." CHE Research Paper 22. Centre for Health Economics(University of York, UK).

Hole, A. R. (2008). "Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment." J Health Econ **27**(4): 1078-1094.

Howard, K. and G. Salkeld (2009). "Does Attribute Framing in Discrete Choice Experiments Influence Willingness to Pay? Results from a Discrete Choice Experiment in Screening for Colorectal Cancer." Value in Health **12**(2): 354-363.

Johnson, F. R., S. Ozdemir and K. A. Phillips (2010). "Effects of simplifying choice tasks on estimates of taste heterogeneity in stated-choice surveys." Soc Sci Med **70**(2): 183-190.

Kjaer, T. and D. Gyrd-Hansen (2008). "Preference heterogeneity and choice of cardiac rehabilitation program: Results from a discrete choice experiment." Health Policy **85**(1): 124-132.

Kruijshaar, M. E., M. L. Essink-Bot, B. Donkers, C. W. Looman, P. D. Siersema and E. W. Steyerberg (2009). "A labelled discrete choice experiment adds realism to the choices presented: preferences for surveillance tests for Barrett esophagus." BMC Med Res Methodol **9**: 31.

Lancsar, E. J., J. P. Hall, M. King, P. Kenny, J. J. Louviere, D. G. Fiebig, I. Hossain, F. C. K. Thien, H. K. Reddel and C. R. Jenkins (2007). "Using discrete choice experiments to investigate subject preferences for preventive asthma medication." Respirology **12**(1): 127-136.

McNamara, A., G. Chen, S. George, R. Walker and J. Ratcliffe (2013). "What factors influence older people in the decision to relinquish their driver's licence? A discrete choice experiment." Accid Anal Prev **55**: 178-184.

Meijer, E. and J. Rouwendal (2006). "Measuring welfare effects in models with random coefficients." Journal of Applied Econometrics **21**(2): 227-244.

Mentzakis, E., M. Ryan and P. McNamee (2011). "Using Discrete Choice Experiments to Value Informal Care Tasks: Exploring Preference Heterogeneity." Health Econ **20**(8): 930-944.

Mentzakis, E., P. Stefanowska and J. Hurley (2011). "A discrete choice experiment investigating preferences for funding drugs used to treat orphan diseases: an exploratory study." Health Econ Policy Law **6**(3): 405-433.

Mohamed, A. F., J. D. Epstein and J. M. Li-McLeod (2011). "Patient and parent preferences for haemophilia A treatments." Haemophilia **17**(2): 209-214.

Naik-Panvelkar, P., C. Armour, J. M. Rose and B. Saini (2012). "Patient preferences for community pharmacy asthma services: a discrete choice experiment." Pharmacoeconomics **30**(10): 961-976.

Negrin, M. A., J. Pinilla and C. J. Leon (2008). "Willingness to pay for alternative policies for patients with Alzheimer's Disease." Health Economics Policy and Law **3**(3): 257-275.

Oteng, B., F. Marra, L. D. Lynd, G. Ogilvie, D. Patrick and C. A. Marra (2011). "Evaluating societal preferences for human papillomavirus vaccine and cervical smear test screening programme." Sex Transm Infect **87**(1): 52-57.

Ozdemir, S., F. R. Johnson and A. B. Hauber (2009). "Hypothetical bias, cheap talk, and stated willingness to pay for health care." J Health Econ **28**(4): 894-901.

Park, M. H., C. Jo, E. Y. Bae and E. K. Lee (2012). "A comparison of preferences of targeted therapy for metastatic renal cell carcinoma between the patient group and health care professional group in South Korea." Value Health **15**(6): 933-939.

Potoglou, D., P. Burge, T. Flynn, A. Netten, J. Malley, J. Forder and J. E. Brazier (2011). "Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data." Soc Sci Med **72**(10): 1717-1727.

Rabe-Hesketh, S., A. Skrondal and A. Pickles (2004). "GLLAMM Manual. U.C. Berkeley *Division of Biostatistics* Working Paper Series." **Paper 160**(University of California, Berkeley).

Regier, D. A., J. M. Friedman, N. Makela, M. Ryan and C. A. Marra (2009). "Valuing the benefit of diagnostic testing for genetic causes of idiopathic developmental disability: willingness to pay from families of affected children." Clin Genet **75**(6): 514-521.

Rennie, L., T. Porteous and M. Ryan (2012). "Preferences for managing symptoms of differing severity: a discrete choice experiment." Value Health **15**(8): 1069-1076.

Revelt, D. and K. Train (1998). "Mixed logit with repeated choices: Households' choices of appliance efficiency level." Review of Economics and Statistics **80**(4): 647-657.

Robyn, P. J., T. Barnighausen, A. Souares, G. Savadogo, B. Bicaba, A. Sie and R. Sauerborn (2012). "Health worker preferences for community-based health insurance payment mechanisms: a discrete choice experiment." BMC Health Serv Res **12**: 159.

Scalone, L., V. Watson, M. Ryan, N. Kotsopoulos and R. Patel (2011). "Evaluation of patients' preferences for genital herpes treatment." Sexually Transmitted Diseases **38**(9): 802-807.

Scuffham, P. A., J. A. Whitty, M. Taylor and R. C. Saxby (2010). "Health system choice: a pilot discrete-choice experiment eliciting the preferences of British and Australian citizens." Appl Health Econ Health Policy **8**(2): 89-97.

Stolk, E. A., M. Oppe, L. Scalone and P. F. Krabbe (2010). "Discrete choice modeling for the quantification of health states: the case of the EQ-5D." Value Health **13**(8): 1005-1013.

Sweeting, K. R., J. A. Whitty, P. A. Scuffham and M. J. Yelland (2011). "Patient preferences for treatment of achilles tendon pain: results from a discrete-choice experiment." Patient **4**(1): 45-54.

Swinburn, P., A. Lloyd, S. Ali, N. Hashmi, D. Newal and H. Najib (2011). "Preferences for antimuscarinic therapy for overactive bladder." BJU Int **108**(6): 868-873.

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

van Helvoort-Postulart, D., T. van der Weijden, B. G. Dellaert, M. de Kok, M. F. von Meyenfeldt and C. D. Dirksen (2009). "Investigating the complementary value of discrete choice experiments for the evaluation of barriers and facilitators in implementation research: a questionnaire survey." Implement Sci **4**: 10.

Waschbusch, D. A., C. E. Cunningham, W. E. Pelham, H. L. Rimas, A. R. Greiner, E. M. Gnagy, J. Waxmonsky, G. A. Fabiano, J. A. Robb, L. Burrows-Maclean, M. Scime and M. T. Hoffman (2011). "A discrete choice conjoint experiment to evaluate parent preferences for treatment of young, medication naive children with ADHD." J Clin Child Adolesc Psychol **40**(4): 546-561.

Whitty, J. A., S. R. Rundle-Thiele and P. A. Scuffham (2012). "Insights from triangulation of two purchase choice elicitation methods to predict social decision making in healthcare." Appl Health Econ Health Policy **10**(2): 113-126.

Whitty, J. A., P. A. Scuffham and S. R. Rundle-Thiele (2011). "Public and decision maker stated preferences for pharmaceutical subsidy decisions: a pilot study." Appl Health Econ Health Policy **9**(2): 73-79.

Whitty, J. A., S. Stewart, M. J. Carrington, A. Calderone, T. Marwick, J. D. Horowitz, H. Krum, P. M. Davidson, P. S. Macdonald, C. Reid and P. A. Scuffham (2013). "Patient preferences and willingness-to-pay for a home or clinic based program of chronic heart failure management: findings from the Which? trial." PLoS One **8**(3): e58347.

Wittink, M. N., M. Cary, T. Tenhave, J. Baron and J. J. Gallo (2010). "Towards Patient-Centered Care for Depression: Conjoint Methods to Tailor Treatment Based on Preferences." Patient **3**(3): 145-157.

Wooldridge, J. M. (2002). "Economic analysis of cross section and panel data." MIT press.

Chapter 8: Whose utility is it anyway? Respondent quality of life and choice experiment preferences, under a veil of altruism.

1. Introduction.

As indicated in chapter 1 (section E1.4), the DCEs that have been deployed in health (see papers reviewed in section D of chapter 1) tend to assume that respondents have a neo-classical self interested perspective when they answer a DCE. However, the DCE analysis for the renal transplant project (chapters 5, 6, 7, and 8) was applied in a context in which altruistic motivations may apply. Indeed, as pointed out in chapter 1, altruistic concerns for the welfare of others have been widely cited as the main motivation for organ donation (Siminoff, Mercer et al. 2007; Gill and Lowes 2008; Patel, Chadha et al. 2011) Admittedly, because we are looking at a slightly different issue from a different direction (i.e. we are considering what criteria should be used to allocate kidneys that others have donated), altruistic considerations might be less pronounced but may still be relevant. This is why I considered it relevant to attempt to establish what motivates patients (i.e. how altruistic they might be), when they express their preferences in terms of who should be prioritized for kidney transplants. Moreover, having attempted to obtain patients' self reported information about whether they were altruistic when answering the DCE renal transplant questionnaire, we then went on to see how the perspective respondents stated they had adopted might have affected their preferences.

This chapter reports in more detail the results obtained from analysis of renal patient responses to our renal transplant DCE questionnaire. As indicated in earlier chapters respondents faced a series of choices which required them to allocate a transplant to either patient A or patient B. Levels of attributes varied between the patient A and patient B choices.

Attributes for the final questionnaire included time spent awaiting a transplant; quality of donor / recipient tissue match; number of adult / child dependents that recipients had; whether recipients had diseases affecting life expectancy; and whether recipients had other diseases affecting quality of life.

For the patient group of respondents, information on respondent quality of life (measured using both Eq-5d and the Visual Analogue Scale) was obtained. We also acquired information upon respondents' self reported altruistic motivations. To this end we posed a question to patients about what perspective they adopted when answering the questionnaire. Response options for this question were either related to answering in terms of what was best for them; answering in terms of what was best for themselves and others; or answering only in terms of what was best for others.

Using econometric analysis the possible impact of altruistic motivations upon a respondent's choice was first assessed using an econometric model which attempted to compare the preferences of respondents who claimed they had only considered others (which I will call 'altruistic responders') with other responders. It should be noted that although we would have liked to have compared responses econometrically for all 3 groups of respondents (i.e. those who claimed they answered in terms of what was best for them, for themselves and others, or only in terms of what was best for others) this was not possible. This was because only 30 / 895 patients (3.3% who indicated the perspective they adopted), said they only considered what was best for them (i.e. that they were purely self-motivated). This provided too small a sample to allow us to establish DCE preferences separately for this group of respondents using interaction dummy variables.

We hypothesized that, if respondents behaved in a self-interested manner, then their own quality of life (as measured by Eq-5d or VAS) might influence their preferences, such that they would favour prioritizing people with similar characteristics to themselves. This would imply that respondents with poor quality of life ratings might be more inclined to prioritize hypothetical transplant recipients with serious diseases affecting life expectancy or with serious diseases affecting quality of life (because both states are likely to be associated with lower scores as measured by Eq-5d or VAS). In contrast, amongst respondents with more altruistic motivations, we might expect less association between their actual quality of life and their willingness to prioritize those with diseases affecting length of life or quality of life.

This hypothesis was examined econometrically. We endeavoured to divide the sample of all the patients who answered a question about the perspective they adopted when answering up into quartiles in terms of respondents quality of life status both as measured by Eq-5d; and also as measured by VAS. We then ran a series of regression analyses to establish whether preferences for attributes differed, comparing the preferences of patients in the upper and lower ranges of quality of life status to those whose reported quality of life was in the mid-range category. We wanted to establish links between respondent quality of life (as measured by Eq-5d), depending on whether respondents were altruistic, and their preferences. So, this analysis was run on a pooled sample; a respondent grouping who admitted to at least some self-interested motivations when answering the DCE questions; and a respondent group who reported only altruistic motivations. We repeated exactly the same analysis with respect to the same three groups to establish the links between quality of life (as measured by VAS), and preferences (i.e. we ran this analysis upon a pooled sample; upon a respondent grouping who admitted to at least some self-interested motivations when answering the DCE questions; and upon a respondent group who admitted to only altruistic motivations).

2. Background.

In chapter 1 (section E1.4) I suggested that, at a theoretical level, the concept of the 'caring externality' (Culyer 1976) might be used to explain altruistic motivations. The concept of the 'caring externality' (Culyer 1976) allows for the possibility that the utility an individual derives from healthcare provision may not solely derive from the utility obtained from treatment or care she / he benefits from. Rather instead altruistic concerns (about the health status of others) might enter into individual's utility functions. The concept of the caring externality, has been used to explain the existence of the NHS and welfare state more generally (Culyer 1980). However, it could also be used to explain why altruistic motivations might apply when patients express their preferences in terms of who should be prioritized for renal transplants.

It should be noted that the concept of the 'caring externality' does not require that decisions need to be completely altruistic (disregarding self interest). The concept implies that individuals will choose on the basis of their own preferences, but their preferences include benevolent ones, such that utility is derived from the amount of utility enjoyed by others (Dowie 1985). There is some evidence (Jacobsson, Carstensen et al. 2005) who conducted a mainstream WTP analysis (which did not involve the use of DCEs) that 'caring externalities' exist. The preferences of 180 Swedish respondents (who primarily worked in healthcare, but others were administrators, farmers, entrepreneurs, students, retired persons and blue-collar workers) were surveyed. Respondents were asked to value 7 different health states with different levels of severity of ill health. For internal preference WTP questions, respondents were asked to reveal their WTP to be cured from each health state. Alternatively the 'caring externality' questions dealt with respondents WTP for someone else to be cured from the same health state. The findings suggested that "Caring externalities were present in all health states. Even concerning the mildest health state, 52% of respondents were willing to pay to

cure the sick person." (Jacobsson, Carstensen et al. 2005, P176). However, the mean value of caring externality WTP was always lower than the mean value of internal WTP, suggesting that respondents consistently valued health benefits to themselves more than health benefits to others. Moreover, respondents were more likely to positively value (express a WTP) for health benefits conferred upon others for more severe health states. This research suggests to me that in the context of a severe condition like renal disease which is severe enough for someone to be on the kidney transplant list, that 'caring externalities' might account for a substantive amount of DCE respondents' utility. Relatively few analyses seem to have looked at the association between results from quality of life measures and preferences findings from choice experiments, although, some work involving both techniques has been undertaken (Bryan, Weatherburn et al. 2001; Ratcliffe, Buxton et al. 2005; Hjelmgren and Anell 2007; Ossa, Briggs et al. 2007; Aspinall, Johnson et al. 2008; Grant, Wileman et al. 2008). I think this is an important issue, and the distinguishing characteristic of the analysis in this chapter, is that we focus upon how quality of life status as measured by Eq-5d and VAS, might impact upon stated preferences expressed in choice experiments, and in particular upon how these links might be affected if you have altruistic responders.

We wanted to establish whether any link between respondent quality of life and a respondent's valuation of attributes might be affected by the perspective respondents adopted when answering the questionnaire. Even if respondents are unaffected by the issues raised, they may be concerned about the impact on the health of others, i.e. they have altruistic concerns (Johansson 1994). Moreover, the very fact that people consent for an organ to be donated without a reward, is difficult to explain if one begins with the assumption that narrow self-interest must explain behaviour (Schenk 1987). Therefore given that altruistic motives do apply in the context of transplantation, it is useful to consider how they might impact upon respondent's stated preferences for prioritizing people for transplants.

Of course altruism may be impure, and motivated by a “warm glow” feeling (Khalil 2004), which could be obtained when making choices that might benefit others. There is also the concept of ‘reciprocal altruism’, whereby people reciprocate, and are motivated by the prospect of future rewards (Trivers 1971; Grossmann 2002). This means that a respondent’s willingness to consider others may be motivated by the consideration that they would like their interests also to be considered by others. However, there is every reason to assume that behaviour based upon self-interest cannot explain all empirical behaviour (Haltiwanger and Waldman 1993), and if we impose such assumptions of behaviour without question, this may fly in the face of empirical observation (Sen 1977).

We therefore wanted to establish what perspective respondents adopted when answering the questionnaire. We asked respondents whether when answering the questionnaire they answered in terms of what would be best for them; in terms of what would be best for them and others; or disregarding what is best for them and only considering what is best for others. The first group of respondents who claimed they considered just what was best for them, have preferences in line with theory which implies that individuals are self interested utility maximisers. The group of respondents who claim they considered what is best for them and others have preferences in line with the concept of the ‘caring externality.’ Clearly only the final group of respondents could be defined as having entirely altruistic motivations.

In this chapter I consider whether respondent’s own quality of life characteristics (as measured by Eq-5d and the Visual Analogue Scale [VAS]) are linked to their preferences for particular attributes. We address this issue using a pooled sample (comprising all the patients who answered the question about the perspective they adopted when answering the questionnaire, and also provided quality of life information). We also run separate

econometric models for responders who indicate their motivations are at least in part self-interested, and altruistic responders respectively, to establish whether a veil of altruism might affect the link between respondent quality of life status and preferences.

Logically we might expect that respondents' own quality of life would impact more upon their preferences if they are less altruistic. This fits with an approach to research in the context of altruism that imputes a motive to an actual choice (Hudson and Jones 1994). Therefore, I would posit that those who have a good quality of life may prefer prioritizing patients with no or less serious diseases affecting quality of life and / or life expectancy, because they themselves fall into this category. Therefore, such people would themselves be more likely to benefit from these priority criteria. In contrast if respondents are altruistic, we would not expect them to be prioritizing people based upon such motivations. This paper assesses the extent to which the perspective that a respondent adopts impacts upon their choice, and how this might affect the link between quality of life and preferences.

3. Methods.

3.1 Piloting the questionnaire

Details are provided in chapter 5 Section 3.1.

3.2 Selection of final attributes and levels

Details of the approach used to select final attributes and levels, are provided in chapter 5 Section 3.2. For convenience the table (table 1 below) which details attributes and levels and associated variable names is reproduced below.

Table 1: Final attributes and levels.

Attribute	Variable name	Levels	Interpretation of coefficients.
Time spent awaiting transplantation	Wait	1 month, 2 years, and 10 years.	Indirect utility of each 1 year reduction in transplant recipient waiting time.
Tissue type matching	Tiss	Non-favourable match: 86% average kidney survival rate post-transplant. Favourable match: 89% average kidney survival rate post-transplant. Perfect match: 90% average kidney survival rate post-transplant.	Indirect utility of prioritizing people for each 1% improvement in kidney survival.
How many child or adult dependents recipients have	Dep	None, 1, or 4 dependents.	Indirect utility of each additional dependent.
Recipient age	Age	20 years, 45 years, and 65 years	Indirect utility for each 1 year reduction in recipient age.
Diseases predominantly affecting life expectancy	dis1	No disease affecting life expectancy (other than Kidney disease) vs. moderate disease (uncontrolled hypertension or obesity) & Kidney disease.	Indirect utility of having no rather than moderate disease predominantly affecting life expectancy.
	dis2	Moderate disease (uncontrolled hypertension or obesity) affecting life expectancy vs. severe disease (heart attack, stroke, or diabetes with complications).	Indirect utility of having moderate disease rather than severe disease predominantly affecting life expectancy.
Diseases predominantly affecting quality of life	ill1	No disease affecting quality of life (other than Kidney disease) vs. moderate disease (mild asthma).	Indirect utility of having no disease rather than a moderate disease predominantly affecting quality of life.
	ill2	Moderate disease (mild asthma) affecting quality of life vs. severe disease (severe arthritis).	Indirect utility of having a moderate disease rather than a severe disease predominantly affecting quality of life.

3.3 Development of the final questionnaire

Details are provided in chapter 5 Section 3.3

3.4 Distribution of the questionnaire

Details are provided in chapter 5 Section 3.4

3.5. Information relevant to this analysis obtained using the patient DCE questionnaire.

Readers should be aware that patient renal transplant questionnaires (see appendix E) also contained both Eq-5d quality of life questions, and the VAS scale. Moreover, of particular importance for this analysis was that a question was posed to help to ascertain how altruistic respondents might have been, when they answered the DCE questionnaire. The question asked:

What perspective did you adopt when answering this questionnaire?: ***(Please tick 1 box only):***

Answering the questions in terms of what would be best for me ❑

Answering the questions in terms of what would be best for me and others ❑

Disregarding what is best for me and only considering what is best for others ❑

3.6 Data analysis.

We use model type A, to compare preferences patient preferences for altruistic patients with respondents who do not claim to only consider others (i.e. who acknowledge that self-interest influence either partly or wholly influences their DCE choices) when making choices. Here Y_{ij} is the binary dependent variable, from individuals i = 1…m, for observations j = 1…n_i. The number of observations n_i varies because the i individuals do not all complete every pairwise choice (a minority of respondents don't answer all choices). The term μ_i is the random effects error term (which allows for multiple responses from i respondents), and ε_{ij} is the standard Probit error term for individuals i for j observations. The other variables are as defined in table 1. The D_A prefix indicates a dummy variable, for the altruistic responders question (see table 2).

$$Y_{ij}= \beta_0+ \beta_1 wait_{ij} +\beta_2 tiss_{ij}+\beta_3 dep_{ij}+\beta_4 age_{ij}+\beta_5 dis1_{ij}+\beta_6 dis2_{ij}+\beta_7 ill1_{ij}+\beta_8 ill2_{ij}$$
$$+ \beta_9 D_A+\beta_{10} D_{Aij} wait_{ij}+\beta_{11} D_{Aij} tiss_{ij}+\beta_{12} D_{Aij} dep_{ij}+\beta_{13} D_{Aij} age_{ij}+\beta_{14} D_{Aij} dis1_{ij}$$
$$+\beta_{15} D_{Aij} dis2+\beta_{16} D_{Aij} ill1_{ij}+\beta_{17} D_{Aij} ill2_{ij}+\mu_i+\varepsilon_{ij}$$

(Model type A: models 1).

Table 2: Dummy variables.

Model type A	**(Model 1).**	
Respondent grouping	**Dummy variable (prefix upon variable name)**	**Coding for dummy variable**
Base group (preferences for non-altruistic patients)	Not / Applicable	Not required: Preferences for non-altruistic responders
Preferences for altruistic patients	D_A	D_A = 1 for altruistic responders. D_A= 0 otherwise
Model type B	**(Models 2,3,4,5,6 and 7).**	
Respondent grouping	**Dummy variable (prefix upon variable name)**	**Coding for dummy variable**
Base group (preferences for patients not in the defined sub-groups)	Not / Applicable	Not required: Preferences of those not in the upper or lower groupings
Those in the lowest quartile of the continuously distributed variable	D_{LQ}	D_{LQ} = 1 for respondents in the lowest quartile D_{LQ} = 0 otherwise
Those in the highest quartile (or QALY =1 category for Eq-5d) of the continuously distributed variable	D_{HQ}	D_{HQ} = 1 for respondents in the highest quartile (or top 28% for which QALY =1 for Eq-5d) D_{HQ} = 0 otherwise

Model type B, can be used when you want to compare the upper and lower quartiles of a continuously distributed patient attribute, such as quality of life, with the inter-quartile range or 'mid-range' grouping. Note we found that 28% of respondents overall had an Eq-5d score of 1. Therefore instead of having an upper quartile when analysing Eq-5d scores, we had an upper quality of life grouping for Eq-5d scores containing all those with an Eq-5d score of 1 (i.e. the top 28%)

For model B, Y_{ij}, μ_i, and ε_{ij} are as previously defined, whilst the D_{LQ} and D_{HQ} prefixes are as defined in table 3.

$$Y_{ij} = \beta_{0+} \beta_1 wait_{ij} +\beta_2 tiss_{ij} +\beta_3 dep_{ij} +\beta_4 age_{ij} +\beta_5 dis1_{ij}+\beta_6 dis2_{ij}+\beta_7 ill1_{ij}+\beta_8 ill2_{ij}+\beta_9 D_{LQ}+\beta_{10} D_{LQij} wait_{ij}$$

$$+\beta_{11} D_{LQij} tiss_{ij}+\beta_{12} D_{LQij} dep_{ij}+\beta_{13} D_{LQij} age_{ij}+\beta_{14} D_{LQij} dis1_{ij} +\beta_{15} D_{LQij} dis2_{ij}+\beta_{16} D_{LQij} ill1_{ij}+\beta_{17} D_{LQij} ill2_{ij}$$

$$+\beta_{18} D_{HQij}+ \beta_{19} D_{HQij} wait_{ij}+\beta_{20} D_{HQij} tiss_{ij}+\beta_{21} D_{HQij} dep_{ij}+\beta_{22} D_{HQij} age_{ij}+\beta_{23} D_{HQij} dis1_{ij}+\beta_{24} D_{HQij} dis2_{ij}$$

$$+\beta_{25} D_{HQij} ill1_{ij}+\beta_{26} D_{HQij} ill2_{ij}+\mu_i+\varepsilon_{ij}$$

(Model type B: models 2,3,4,5, 6 and 7).

This model can be applied to establish how respondent Eq-5d status, or VAS status affects their preferences, both for a pooled sample; a respondent group who do not claim to only consider others (who are at least in part self-interested); or altruistic responders (who consider only others).

3.7 Establishing Marginal Rate of Substitution (MRS).

To express the value of changes in attributes with respect to changes in another we calculate MRS (see table 3). MRS values changes in the other variables compared with a 1 year change in waiting time. To establish whether MRS is significant we used the Delta method (Wooldridge 2002) which was executed using the 'nlcom' command in STATA v. 9.2, to derive 95% confidence intervals. We use Wald tests (executed using 'testnl' in STATA) to establish whether there is a statistically significant difference in MRS between the base group and dummy variables group(s). Difference in MRS is indicated by a Wald test p-value ≤ 0.05.

Table 3: Calculating MRS.

Model type A **Variable**	**Base group MRS**	**Model type A** **Variable**	**Defined sub-group MRS**
wait	N / A	**wait**	N / A
tiss	β_2 / β_1	**tiss**	$(\beta_2 + \beta_{11}) / (\beta_1 + \beta_{10})$
dep	β_3 / β_1	**dep**	$(\beta_3 + \beta_{12}) / (\beta_1 + \beta_{10})$
age	β_4 / β_1	**age**	$(\beta_4 + \beta_{13}) / (\beta_1 + \beta_{10})$
dis1	β_5 / β_1	**dis1**	$(\beta_5 + \beta_{14}) / (\beta_1 + \beta_{10})$
dis2	β_6 / β_1	**dis2**	$(\beta_6 + \beta_{15}) / (\beta_1 + \beta_{10})$
ill1	β_7 / β_1	**ill1**	$(\beta_7 + \beta_{16}) / (\beta_1 + \beta_{10})$
ill2	β_8 / β_1	**ill2**	$(\beta_8 + \beta_{17}) / (\beta_1 + \beta_{10})$
Model type B **Variable**	**Base group MRS**	**Model type B** **Variable**	**Lower quartile MRS**
wait	N / A	**wait**	N / A
tiss	β_2 / β_1	**tiss**	$(\beta_2 + \beta_{11}) / (\beta_1 + \beta_{10})$
dep	β_3 / β_1	**dep**	$(\beta_3 + \beta_{12}) / (\beta_1 + \beta_{10})$
age	β_4 / β_1	**age**	$(\beta_4 + \beta_{13}) / (\beta_1 + \beta_{10})$
dis1	β_5 / β_1	**dis1**	$(\beta_5 + \beta_{14}) / (\beta_1 + \beta_{10})$
dis2	β_6 / β_1	**dis2**	$(\beta_6 + \beta_{15}) / (\beta_1 + \beta_{10})$
ill1	β_7 / β_1	**ill1**	$(\beta_7 + \beta_{16}) / (\beta_1 + \beta_{10})$
ill2	β_8 / β_1	**ill2**	$(\beta_8 + \beta_{17}) / (\beta_1 + \beta_{10})$
Model type C **Variable**	**Higher quartile MRS**		
wait			
tiss	$(\beta_2 + \beta_{20}) / (\beta_1 + \beta_{19})$		
dep	$(\beta_3 + \beta_{21}) / (\beta_1 + \beta_{19})$		
age	$(\beta_4 + \beta_{22}) / (\beta_1 + \beta_{19})$		
dis1	$(\beta_5 + \beta_{23}) / (\beta_1 + \beta_{19})$		
dis2	$(\beta_6 + \beta_{24}) / (\beta_1 + \beta_{19})$		
ill1	$(\beta_7 + \beta_{25}) / (\beta_1 + \beta_{19})$		
ill2	$(\beta_8 + \beta_{26}) / (\beta_1 + \beta_{19})$		

4. Results.

4.1. Results - Sample characteristics.

Table 4: Sample characteristics.

	Patients (n = 908)
AGE	
Mean age	54.88 years
GENDER	
Male	508 (55.9%)
Female	397 (43.7%)
Not indicated	3 (0.3%)
ETHNICITY	
White (British)	799 (88%)
White ethnic minorities	27 (2.9%)
Non-white ethnic minorities	69 (7.6%)
Asian groups (also included in non-Asian ethnic minority category)	50 (5.5%)
Not indicated	13 (1.4%)
DEPENDENT CHILDREN	
0	755 (83.1%)
1	72 (7.9%)
2	49 (5.4%)
3	12 (1.3%)
> 3	7 (0.8%)
Not indicated	13 (1.4%)
DEPENDENT ADULTS	
0	750 (82.6%)
1	121 (13.3%)
2	17 (1.9%)
> 2	8 (0.9%)
Not indicated	12 (1.3%)

Table 4 indicates respondent characteristics. UK Renal Registry data (Byrne, Ford et al. 2008; Byrne, Steenkamp et al. 2008) was used to assess patient sample representativeness. Of the 895/ 908 patients indicating ethnicity, 799 / 895 patients (89.3%) were white (British), and 27 / 895 (3%) were white ethnic minorities, so 92.3% are white. UK incidence data

(Byrne, Ford et al. 2008) suggested 79.7% of renal patients are white, so whites are over-represented in our survey. Overall, 69 / 895 (7.7%) patients indicating ethnicity were non-white, compared with a 20.3% incidence rate (Byrne, Ford et al. 2008), 50 / 69 non-white patients were South Asians (5.6% of those indicating ethnicity) compared to a 10.5% incidence (Byrne, Ford et al. 2008). 508 /908 patients (55.9%) were male, 397 / 908 (43.7%) were female, 3 / 908 (0.3%) did not say. Graphically presented Renal Registry data (Byrne, Ford et al. 2008) reassuringly indicated slightly higher proportions of male than female patients across age groups. Average sample patient age was 54.88 years (median 57 years), and Renal Registry data median age (57.3 years) was virtually identical (Byrne, Steenkamp et al. 2008).

The patient sample comprised: 468 / 908 (51.5%) with successful transplants; 118 / 908 (13%) whose transplant failed; 279 / 908 (30.7%) awaiting transplants (average wait 22.6 months). Some patients whose transplant failed also appeared as awaiting transplantation; 237 / 908 (26.3%) had dialysis without transplantation; and 57 / 908 (6.3%) had kidney disease, not requiring dialysis. Renal Registry prevalence data (Byrne, Steenkamp et al. 2008) suggests 46.9% of patients have successful transplants (close to our figure). There is no incidence / prevalence data for other patient categories. Amongst non-whites (including Asians) figures are 18 / 69 patients (26%) with successful transplants; 10 / 69 (14.5%) whose transplant failed; 35 / 69 patients (50.7%) awaiting a transplant on dialysis (average wait: 21.45 months); and 3 / 69 (4.3%) with kidney disease, not requiring dialysis. Unfortunately renal registry data (Byrne, Steenkamp et al. 2008) does not indicate ethnicity. However, lower percentage figures for successes, and higher transplant failures figures are expected (ethnic minorities are less likely to be closely matched).

Overall 895 / 908 patients responded to the question about the perspective they adopted, 30 considered only what was best for them, 323 claimed they considered what was best for them and others, and 542 considered what was best only for others (altruistic responders).

4.2. Econometric results.

Table 5 - Model 1: Whole sample – Other patients vs. Altruistic patients

Attribute	Coefficient for patients in general	Implied MRS for patients in general	Coefficient for dummy variable for the patients who are Altruistic.	Implied MRS for the altruistic group	Wald test p-values
wait	.0478**	1	-.0057	1	
tiss	.0568**	1.19** (0.71 / 1.66)	.0101	1.59** (1.12 / 2.05)	p=0.0490
dep	.0830**	1.74** (1.34 / 2.14)	-.0326**	1.20** (0.85 / 1.54)	p=0.0000
age	.0081**	0.17** (0.12 / 0.22)	-.0019	0.15** (0.10 / 0.20)	p=0.0001
dis1	.0458	0.96 (-0.55 / 2.46)	-.0750	-0.70 (-2.09 / 0.70)	p=0.1257
dis2	.6181**	12.92** (10.41 / 15.44)	.1087	17.26** 14.53 / 20.00)	p=0.0000
ill1	-.1032*	-2.16* (-0.28 / -4.04)	-.0294	-3.15** (-1.41 / -4.89)	p=0.4912
ill2	.2158**	4.51** (2.96 / 6.07)	-.0510	3.91** (2.47 / 5.36)	p=0.0015
Intercepts	.1315**		-.0148		
% of actual values predicted:	63.14%	**Sample:**	908 patients (542 are altruistic)	**McFaddens R^2:**	0.112
LR test (λ):	23.49	**Jointly significant?**	Yes: CV for 9 dfs = 16.92	**Log-likelihood:**	-4989.99

4.3. Comparing those who only thought of others (labelled 'altruistic responders') vs. other respondents.

I used an interaction dummy variable model to establish whether preferences varied for those claiming to have only thought of others (who I labelled 'altruistic' vs. other patients (model 1, table 5). The model accurately predicts 63.14% of responses and McFaddens $R^2 = 0.112$ The findings of a likelihood ratio (LR) test for joint significance ($\lambda = 23.49$, when the critical value (CV) for 9 degrees of freedom is 16.92) suggest that the dummy variables for the altruistic responder group are jointly significant. However, 7 / 8 of the dummy interaction variables are not statistically significant. Findings suggest that those who identify themselves as altruistic responders attach a lower valuation to prioritising patients having child / adult dependents, and there is also a statistically significant difference in MRS with respect to this variable (p=0.000) according the Wald test. The MRS (MRS = 1.20 [CI: 0.85 / 1.54] vs. 1.74 [CI: 1.34 / 2.14]) for this variable is lower for altruistic responders than for non-altruistic responders. Wald tests also suggest that MRS is statistically significantly different with respect to 4 other variables. The variable tiss varies (p=0.0490) according to the Wald test. MRS (MRS = 1.59 [CI: 1.12 / 2.05] vs. 1.19 [CI: 0.71 / 1.66]) is higher amongst altruistic responders, who value prioritizing those with better tissue matches more highly. MRS for age is also significantly different (p=0.0001), and MRS (MRS = 0.15 [CI: 0.10 / 0.20] vs. 0.17 [CI: 0.12 / 0.22]), is lower for altruistic responders who value prioritizing younger recipients less. MRS for dis1 also differs (p=0.0000), and MRS (MRS =17.26 [CI: 14.53 / 20.00] vs. 12.92 [CI: 10.41 / 15.44]) is higher amongst altruistic responders, who value prioritizing those with moderate rather than severe diseases affecting life expectancy more highly. Finally there is evidence that MRS for ill2 (p=0.0015) differs between the 2 groups. The MRS (MRS = 3.91 [CI: 2.47 / 5.36] vs. 4.51 [CI: 2.96 / 6.07]), is lower for altruistic responders suggesting they would prioritize those with moderate rather than severe diseases affecting life expectancy

less than non-altruistic respondents. These results taken together suggest that there may be a case for modelling the preferences of non-altruistic and altruistic respondents separately.

4.4. Analysis of data for the <u>whole</u> sample. Comparing the preferences of those with preferences in the lower and upper range of Eq-5d scores, with mid-range scores.

I considered whether patient responses are affected by patient quality of life as measured using Eq-5d. We ran an econometric model with interaction dummy variables both for lower and upper quartiles (model 2 – table 6) for the whole sample of patients who answered the 'altruism' question. In actual fact c.28% of the combined sample of respondents had a QALY score of 1. So to accommodate this, our top grouping only approximates a quartile, being slightly larger, it contains 28% rather than 25% of patient respondents. This means that we have a mid-range sample which contains 47% of respondents i.e. it excludes 28% of respondents which have QALY =1, and excludes the 25% of respondents in the lowest QALY quartile (it is thus in reality marginally less than the inter-quartile range so we refer to it as the 'mid-range' sample).

The model accurately predicts 63.26% of responses and McFaddens $R^2 = 0.113$
Findings from the likelihood ratio test ($\lambda = 20.68$, critical value = 28.87 for 18 degrees of freedom) suggest that the dummy variables are not jointly significant. Moreover, none of the coefficients on any of the dummy variables are statistically significant, so variable impacts do not appear to vary according to which grouping of Eq-5d status respondents are in.

Table 6 – model 2: <u>Whole sample of patients</u> with interaction dummy variables for those in the lowest and highest quartiles for quality of life as measured by Eq-5d.

Attribute	**Coefficient for 'mid range'**	**MRS for 'mid range'**	**Coefficient on dummy for those in quartile 1**	**MRS for lowest Eq-5d Quartile**	**Wald test p-values**
wait	.0466**	1	-.0069	1	
tiss	.0621**	1.33** (0.87 / 1.79)	.0060	1.72** (0.93 / 2.50)	p=0.0237
dep	.0570**	1.22** (0.87 / 1.58)	.0203	1.95** (1.31 / 2.59)	p=0.0727
age	.0081**	0.17** (0.12 / 0.23)	-.0034	0.12** (.04 / 0.20)	p=0.0000
dis1	-.0014	-0.03 (-1.46 / 1.40)	-.0130	-0.36 (-2.68 / 1.96)	p=0.8826
dis2	.7009**	15.03** (12.44 / 17.62)	-.1094	14.91** (10.72 / 19.09)	p=0.0000
ill1	-.0940*	-2.02* (-0.20 / -3.83)	-.0058	-2.52 (-5.38 / 0.35)	p=0.4262
ill2	.1552**	3.33** (1.84 / 4.82)	.0069	4.09** (1.70 / 6.47)	p=0.1041
Intercepts	.1225**		-.0061		
Attribute	**Coefficient for 'mid range'**	**MRS for 'mid range'**	**Coefficient on dummy for the top 28% of scores.**	**MRS for highest Eq-5d Quartile**	**Wald test p-values**
wait	.0466**	1	.0003	1	
tiss	.0621**	1.33** (0.87 / 1.79)	-.0016	1.29** (0.70 / 1.88)	p=0.0095
dep	.0570**	1.22** (0.87 / 1.58)	.0045	1.31** (0.85 / 1.77)	p=0.0083
age	.0081**	0.17** (0.12 / 0.23)	-.0015	0.14** (0.08 / 0.21)	p=0.0005
dis1	-.0014	-0.03 (-1.46 / 1.39)	.0071	0.12 (-1.73 / 1.97)	p=0.9161
dis2	.7009**	15.03** (12.44 / 17.62)	.0313	15.60** (12.18 / 19.02)	p=0.0000
ill1	-.0940*	-2.01* (-0.19 / -3.82)	-.0930	-3.98** (-1.69 / -6.28)	p=0.9749
ill2	.1552**	3.33** (1.83 / 4.81)	.1075	5.60** (3.67 / 7.53)	p=0.6410
Intercepts	.1225**		.0154		
% of actual values predicted:	63.26%	**Sample:**	892 patients: 223 in first quartile and 253 in top 28%.	**McFaddens R^2:**	0.113
LR test (λ):	20.68	**Jointly significant?**	No : CV for 18 dfs = 28.87	**Log-likelihood:**	-4901.1

There is however some evidence that MRS for individual attributes may differ in the lowest and highest quartile groups. For the respondents in the lowest Eq-5d quartile, Wald tests suggest there is a difference in MRS compared to the mid-range with respect to tiss (p=0.0237), age (p=0.0000), and dis2 (p=0.0000). Whilst for those in the QALY = 1 group, there is a statistically significantly difference relative to the mid-range for tiss (p=0.0095), dep (p=0.0083), age (p=0.0005), and dis1 (p=0.0000).

Taking the findings for the lowest quartile group first. For tiss MRS differences (MRS = 1.72 [CI: 0.93 / 2.50] vs. 1.33 [CI: 0.87 / 1.79]), suggest that those in the lowest quartile of Eq-5d status may place more emphasis upon prioritising those who are a good tissue match compared to other patients, this is confirmed by a Wald test (p=0.0237). In relation to age, MRS differs (p=0.0000) and the differences (MRS = 0.12 [CI: 0.04 / 0.20] vs. 0.17 [CI: 0.12 / 0.23]) suggest those in the lowest quartile in terms of quality of life place less emphasis upon prioritizing younger patients. For dis2 the MRS difference (MRS = 14.91 [CI: 10.72 / 19.09] vs. 15.03 [CI: 12.44 / 17.62]) is also confirmed by the Wald test (p=0.0000), but it is very slight.

For the highest Eq-5d grouping (top 28% for which QALY = 1) the MRS for tiss (MRS = 1.29 [CI: 0.70 / 1.88] vs. 1.33 [CI: 0.87 / 1.79]), differs between the highest quartile and the mid-range according to a Wald test (p=0.0095). However, the difference in MRS is slight. With respect to dep (MRS = 1.31 [CI: 0.85 / 1.77] vs. 1.22 [CI: 0.87 / 1.58]), the point estimates suggest that those in the highest 28% of the Eq-5d distribution may prioritise those with child or adult dependents more than those in the mid-range, something confirmed by the Wald test (p=0.0083), but, the difference is marginal. In relation to age (MRS = 0.14 [CI: 0.08 / 0.21] vs. 0.17 [CI: 0.12 / 0.23]) the Wald test suggests that MRS differs for those in the highest Eq-5d

grouping (p=0.0005). However, the difference in MRS is slight, but suggests a lower preference for those with QALY = 1 for prioritizing the young.

For dis2 (MRS = 15.60 [CI: 12.18 / 19.02] vs. 15.03 [CI: 12.44 / 17.62]), and the Wald test again suggests that MRS varies between the groups (p=0.0000). However, once again the difference in relative terms is not large. Wald tests do not confirm any other differences in MRS.

4.5. Analysis of data for (for respondents who do not claim to only consider others). Comparing the preferences of those with preferences in the lower and upper range of Eq-5d scores, with mid-range scores.

You can consider whether it is appropriate to model preferences separately for altruistic vs. non-altruistic respondents for Eq-5d data, or whether a pooled model is adequate by comparing the restricted pooled model (model 2) with the unrestricted separate models for non-altruistic (model 3) and altruistic (model 4) respondents using a likelihood ratio test. The value of λ (42.43) just exceeds the critical value (40.11) at the 5% level, which suggests that using 2 separate models is more appropriate.

In model 3 (table 7) I conduct the same analysis relating to Eq-5d status and preferences for those who are 'non-altruistic' responders. When I refer to the group which is in the top 28% for Eq-5d status or the lowest quartile, I am referring to those in the top 28% for the patient sample overall (they comprise 103 / 359 [28.7%] of the non-altruistic sample).

Table 7 – model 3: Analysis for the group of patients who do not claim to only consider others with interaction dummy variables for those in the lowest and highest quartiles relating to quality of life as measured by Eq-5d.

Attribute	Coefficient for 'mid range'	MRS for 'mid range'	Coefficient on dummy for those in quartile 1	MRS for lowest Eq-5d quartile	Wald test p-values
wait	.0499**	1	-.0029	1	
tiss	.0669**	1.34** (0.64 / 2.04)	-.0261	0.87 (-0.02 / 1.75)	p=0.0188
dep	.0720**	1.44** (0.89 / 2.00)	.0307	2.18** (1.33 / 3.04)	p=0.2091
age	.0086**	0.17** (0.09 / 0.25)	-.0048	0.08 (-0.01 / 0.18)	p=0.0020
dis1	.0154	0.31 (-1.87 / 2.48)	.0918	2.28 (-0.69 / 5.25)	p=0.5551
dis2	.6581**	13.18** (9.52 / 16.84)	-.1680	10.43** (5.96 / 14.89)	p=0.0000
ill1	-.0362	-0.72 (-3.53 / 2.08)	-.1244	-3.42 (-6.98 / 0.15)	p=0.5963
ill2	.1521**	3.05** (0.80 / 5.29)	.0852	5.05** (2.04 / 8.06)	p=0.6658
Intercepts	.1190**		-0025.		
Attribute	**Coefficient for 'mid range'**	**MRS for 'mid range'**	**Coefficient on dummy for patients in the top 28% of Eq-5dscores**	**MRS for patients in the top 28% of Eq-5d scores**	**Wald test p-values**
wait	.0499**	1	-.0002	1	
tiss	.0669**	1.34** (0.64 / 2.04)	-.0133	1.08** (0.19 / 1.97)	p=0.0456
dep	.0720**	1.44** (0.89 / 2.00)	.0092	1.63** (0.89 / 2.37)	p=0.0552
age	.0086**	0.17** (0.09 / 0.25)	.0030	0.23** (0.12 / 0.34)	p=0.2109
dis1	.0154	0.31 (-1.87 / 2.48)	.0112	0.54 (-2.30 / 3.37)	p=0.9789
dis2	.6581**	13.18** (9.52 / 16.84)	.0376	13.98** (9.05 / 18.90)	p=0.0005
ill1	-.0362	-0.72 (-3.53 / 2.08)	-.1292	-3.32** (-6.85 / -0.210)	p=0.5826
ill2	.1521**	3.05** (0.80 / 5.29)	.1719	6.51** (3.50 / 9.52)	p=0.8477
Intercepts	.1190**		.0520		
% of actual values predicted:	63.28%	**Sample:**	359 patients: 103 in 1st quartile and 94 (top 28%)	**McFaddens R^2:**	0.120
LR test (λ):	18.77	**Jointly significant?**	No : CV for 18 dfs = 28.8	**Log-likelihood:**	-.1951.8

Likewise those referred to as in the bottom quartile are those in the bottom 25% of the patient sample overall (they comprise of 94 / 359 [26.2%] of responders who do not claim to only

consider others**)**. I wanted to ensure that those respondents classified as in the top 28% of the sample, or in the bottom quartile of the sample overall, also formed part of those same groups when I ran two separate models in the interests of comparability. I therefore ensured that the Eq-5d cut off points for the highest quality of life group (was always QALY =1), and for the lowest quality of life group (lowest quartile), remained consistent. I adopted the same approach for the later analysis relating to respondent Visual Analogue Scale (VAS) ratings, and their preferences.

Overall 359 / 892 patients formed th**is** respondent group. The model accurately predicts 63.26% of responses and McFaddens R^2 = 0.120.The likelihood ratio test (which is λ = 18.77, compared with a critical value of 28.8 for 18 degrees of freedom), suggests that like the sample overall, the dummy variables are not statistically significant. Furthermore none of the coefficients on dummy variables are statistically significant at the 5% level. This means there is no evidence that attribute values differ in a jointly significant manner between patients for whom QALY = 1, or those in the lowest Eq-5d status quartile, compared with the 'mid range' sample in this group of the sample who have not claimed to only consider others.

There is however evidence that MRS for certain variables might differ significantly. The value of MRS is different for 3 variables (wait, age, and dis2) comparing those in the lowest Eq-5d quartile, and the mid-range part of the sample. There is a difference in MRS for tiss (p=0.0188), the difference in MRS (MRS = 0.87 [CI: -0.02 / 1.75] vs. 1.34 [CI: 0.64 / 2.04]) which indicates that this group of patient respondents in the lowest Eq-5d score grouping have less of a preference for prioritizing people with a better tissue match.

There is also evidence (p=0.0020) that MRS varies for age amongst non-altruistic responders. It appears that those in the lowest Eq-5d quartile for quality of life, do not

necessarily value prioritizing younger recipients (MRS is not significantly different from zero at the 5% level), whilst those in the 'mid-range' of the sample do value prioritizing younger recipients. There is also evidence that MRS for dis2 (p=0.0000). The MRS (MRS = 10.43 [CI: 5.96 / 14.89] vs. 13.18 [CI: 9.52 / 16.84]) is lower for this group of patients in the lowest Eq-5d quartile, compared to the 'mid-range' of the sample. This suggests that for this sample those who are in the lowest Eq-5d quartile value prioritizing those with moderate rather than severe diseases affecting quality of life, less than non-altruists in the 'mid-range' of the quality of life spectrum. It makes sense that responders who do not claim to only consider others who have a low quality of life, should prefer not to prioritize those with moderate rather and severe diseases affecting life expectancy as much as those with mid-range Eq-5d ratings, as it fits with them being at least in part self interested.

Comparing those in the 'mid-range' with those with Eq-5d scores equal to 1 (perfect health) there is evidence that amongst who do not claim to only consider others, MRS varies with respect to two variables. It varies for tiss (p=0.0456), and MRS differences (MRS = 1.08 [CI: 0.19 / 1.97] vs. 1.34 [CI: 0.64 / 2.04]), suggest that amongst this respondent group there is evidence that those in the top 28% (QALY =1), have a slightly lower preference for prioritizing towards those with better recipient / donor tissue matches, compared to those in the Eq-5d 'mid-range'. Also MRS for dis2 differs (p=0.0005). Differences in MRS (MRS = 13.98 [CI: 9.05 / 18.90] vs. 13.18 [CI: 9.52 / 16.84]) suggest that non-altruistic respondents with QALY =1, attach a marginally greater priority towards prioritizing those with moderate rather than severe diseases affecting life expectancy, which can be explained in terms of enlightened self-interest There is no evidence from Wald tests that MRS varies between other dummy variables and the 'mid-range' group.

4.6. Analysis of data for (for respondents who claim to only consider others, who we label 'altruistic'). Comparing the preferences of those with preferences in the lower and upper range of Eq-5d scores, with mid-range scores.

For the altruistic sample (model 4, table 8), once again like models 2, and 3, there is no evidence that dummy variables are jointly significant ($\lambda = 20.33$, against a critical value of 28.8 for 18 degrees of freedom). The model accurately predicts 64.82% of responses and McFaddens $R^2 = 0.114$.Moreover once again none of the interaction dummy variables are significant. Wald tests do however provide some evidence that MRS varies between the groups. Comparing those in the lowest Eq-5d quartile, with the mid-range Eq-5d status group, there is evidence that valuation of age differs (p=0.0037) and MRS (MRS = 0.17 [CI: 0.04 / 0.30] vs. 0.18 [CI: 0.11 / 0.25]) is very marginally lower for those in the lowest Eq-5d status quartile. Similarly for dis2, MRS differs (p=0.0000) and MRS is higher for those in the lowest Eq-5d status grouping (MRS = 20.43 [CI: 12.34 / 28.52] vs. 16.52 [CI: 12.85 / 20.18]). This suggests that altruistic patients in the lowest Eq-5d status quartile would prioritize those with moderate rather than severe diseases affecting life expectancy more than altruistic responders in the mid-range in terms of Eq-5d status. This is a finding that can be explained by them being altruistic.

Table 8 – model 4: <u>Altruistic patients</u> with interaction dummy variables for those in the lowest and highest quartiles relating to quality of life as measured by Eq-5d.

Attribute	Coefficient for 'mid range'	MRS for 'mid range'	Coefficient on dummy for those in quartile 1	MRS for lowest Eq-5d quartile	Wald test p-values
wait	.0443**	1	-.0104	1	
tiss	.0600**	1.36** (0.73 / 1.97)	.0318	2.71** (1.26 / 4.17)	p=0.3959
dep	.0468**	1.06** (0.59 / 1.53)	.0087	1.64** (0.68 / 2.60)	p=0.1467
age	.0079**	0.18** (0.11 / 0.25)	-.0022	0.17** (0.04 / 0.30)	p=0.0037
dis1	-.0116	-0.26 (-2.17 / 1.65)	-.0918	-3.06 (-6.93 / 0.82)	p=0.4386
dis2	.7316**	16.52** (12.85 / 20.18)	-.0408	20.43** (12.34 / 28.52)	p=0.0000
ill1	-.1279*	-2.88* (-5.30 / -0.48)	.0697	-1.72 (-6.35 / 2.90)	p=0.1565
ill2	.1548**	3.50** (1.49 / 5.50)	-.0539	2.98** (0.84 / 6.81)	p=0.0739
Intercepts	.1239**		-.0090		
Attribute	**Coefficient for 'mid range'**	**MRS for 'mid range'**	**Coefficient on dummy for patients in the top 28% of Eq-5d scores**	**MRS for patients in the top 28% of Eq-5d scores**	**Wald test p-values**
wait	.0443**	1	.0021	1	
tiss	.0600**	1.35** (0.73 / 1.97)	.0039	1.38** (0.60 / 2.15)	p=0.0741
dep	.0468**	1.06** (0.59 / 1.53)	.0054	1.12** (0.54 / 1.71)	p=0.1018
age	.0079**	0.18** (0.11 / 0.25)	-.0042	0.08** (0.00 / 0.16)	p=0.0007
dis1	-.0116	-0.26 (-2.17 / 1.65)	-.0016	-0.28 (-2.68 / 2.11)	p=0.9250
dis2	.7316**	16.52** (12.85 / 20.18)	.0292	16.41** (11.85 / 20.96)	p=0.0000
ill1	-.1279*	-2.89* (-5.30 / -0.48)	-.0751	-4.38 (-1.42 / -7.34)	p=0.7221
ill2	.1548**	3.49** (1.49 / 5.50)	.0773	5.00** (2.53 / 7.48)	p=0.5211
Intercepts	.1239**		.0019		
% of actual values predicted:	64.82%	**Sample:**	533 patients: 120 in first quartile and 159 in top 28%	**McFaddens R^2:**	0.114
LR test (λ):	20.33	**Jointly significant?**	No : CV for 18 dfs = 28.8	**Log-likelihood:**	-.2928.09

Comparing those in the top Eq-5d status group (QALY = 1) with those in the mid-range, MRS is statistically different with respect to age (p=0.0007) and dis2 (p=0.0000). MRS for difage (MRS = 0.08 [CI: 0.00 / 0.16] vs. 0.18 [CI: 0.11 / 0.25]) is lower for altruistic responders who have QALY = 1. This suggests that altruists in the highest Eq-5d status grouping would prioritize younger recipients less than altruists with mid-range quality of life. There is also evidence that MRS for dis2 differs between those in the highest Eq-5d grouping compared with the mid-range, according to the Wald test (p=0.0000), however MRS (MRS = 16.41 [CI: 11.85 / 20.96] vs. 16.52 [CI: 12.85 / 20.18]) varies only slightly. There is however evidence that those in highest Eq-5d status groupings value prioritizing those with moderate rather than severe diseases affecting quality of life marginally less, which is again compatible with them being altruistic.

4.7. Analysis of data for the whole sample. Comparing the preferences of those with preferences in 1st and 4th quartiles of VAS scores, with the inter-quartile range.

We assess the impact of differences in quality of life as measured by VAS upon preferences, in model 5 (table 9) for both non-altruistic and altruistic patients. We evaluate whether there are differences between both the lower and upper quartiles in terms of quality of life relative to the inter-quartile range. The model accurately predicts 62.66% of responses and McFaddens R^2 = 0.112.Evidence from the likelihood ratio test suggests that the dummy variables for lower and upper quartiles are jointly significant (λ = 34.82, the critical value for 18 degrees of freedom is 28.87).

Table 9 – model 5: Patients overall: Dummy variables for 1st & 4th VAS quartiles.

Attribute	Coefficient for inter-quartile range	MRS for inter-quartile range	Coefficient on dummy for those in quartile1	MRS for lowest quartile	Wald test p-values
wait	.0449**	1	.0035	1	
tiss	.0728**	1.62** (1.13 / 2.11)	-.0181	1.13** (0.55 / 1.71)	p=0.0003
dep	.0743**	1.65** (1.27 / 2.04)	-.0070	1.39** (0.92 / 1.86)	p=0.0000
age	.0066**	0.15** (0.10 / 0.20)	-.0005	0.13** (0.06 / 0.19)	p=0.0102
dis1	-.0272	-0.61 (-2.07 / 0.86)	.0550	0.57 (-1.28 / 2.43)	p=0.3140
dis2	.733**	16.30** (13.54 / 19.07)	-.216**	10.67** (7.80 / 13.54)	p=0.0000
ill1	-.1344*	-2.99** (-1.17 / -4.81)	.0462	-1.82 (-4.15 / 0.52)	p=0.0865
ill2	.2039**	4.54** (3.03 / 6.04)	-.0432	3.32** (1.37 / 5.26)	p=0.0049
Intercepts	.1446**		-.0352		
Attribute	**Coefficient for inter-quartile range**	**MRS for inter-quartile range**	**Coefficient on dummy for those in quartile 4**	**MRS for highest quartile**	**Wald test p-values**
wait	.0449**	1	-.0056	1	
tiss	.0728**	1.62** (1.13 / 2.11)	-.0242	1.24** (0.50 / 1.97)	p=0.0001
dep	.0743**	1.65** (1.27 / 2.04)	-.0383**	0.91** (0.36 / 1.47)	p=0.0000
age	.0066**	0.15** (0.10 / 0.20)	.0017	0.21** (0.12 / 0.30)	p=0.0585
dis1	-.0272	-0.61 (-2.07 / 0.86)	.0350	0.20 (-2.08 / 2.48)	p=0.4494
dis2	.733**	16.30** (13.54 / 19.07)	.0047	18.74** (13.98 / 23.51)	p=0.0000
ill1	-.1344*	-2.99** (-1.17 / -4.81)	.0134	-3.08** (-0.21 / -5.95)	p=0.1677
ill2	.2039**	4.54** (3.03 / 6.04)	.0336	4.33** (1.94 / 6.72)	p=0.0069
Intercepts	.1446**		-.0491		
% of actual values predicted:	62.66%	**Sample:**	895 patients: 224 in first quartile and 224 in 4th quartile.	**McFaddens R^2:**	0.112
LR test (λ):	34.82	**Jointly significant?**	Yes : CV for 18 dfs = 28.87	**Log-likelihood:**	-4916.88

Only 2 of the interaction dummy variables proved to be significant at the 5% significance level. There is evidence at the 1% significance level that the variable dis2 differs for those in the lowest quartile of VAS scores relative to the inter-quartile range. MRS for dis2 (MRS = 10.67 [CI: 7.80 / 13.54] vs. 16.30 [CI: 13.54 / 19.07]), differs according to the Wald test (p=0.0000). These differences in MRS suggest that those in the lowest quartile of quality of life (as measured by VAS) prioritize those with moderate as opposed to severe diseases affecting life expectancy by less. There is also evidence at the 1% level that dep differs in terms of attribute impacts for those in the highest quartile (the dummy variable on dep is significant at the 1% level). MRS for this variable is also different according to the Wald test (p=0.0000), and MRS is lower for those in the highest quality of life quartile (MRS = 0.91 [CI: 0.36 / 1.47] vs. 1.65 [CI: 1.27 / 2.04]), who value prioritizing those with dependents less.

Evidence from the Wald tests suggests that there are also differences in MRS with respect to a number of variables. Amongst respondents in the lowest quartile of VAS scores, the Wald tests suggest that there are statistically significant differences in MRS with respect to tiss (p=0.0003); dep (p=0.0000); age (p=0.0102); dis2 (p=0.0000) and dif_ill_m_to_s (p=0.0049) compared to the inter-quartile range. Amongst those in the highest quartile in terms of VAS score, the evidence is that according to the Wald test there are statistically significant differences in MRS with respect to tiss (p=0.0001); dep (p=0.0000); dis2 (p=0.0069) compared to the inter-quartile range.

In relation to tiss, comparing the lower quartile VAS score with those not in the inter-quartile range (MRS = 1.13 [CI: 0.55 / 1.71] vs. 1.62 [CI: 1.13 / 2.11]), the Wald test (p=0.0003) suggests that those with lower quartile VAS scores have a lower preference for transplanting, to those with better tissue matches. With respect to dep (MRS = 1.39 [CI: 0.92 / 1.86] vs. 1.65 [CI: 1.27 / 2.04]), the Wald test suggests that MRS varies between the 2 groups (p=0.0000).

Those with lower quartile VAS scores may prioritise those with child or adult dependents less than other respondents.

For age (MRS = (0.13 [CI: 0.06 / 0.19] vs. 0.15 [CI: 0.10 / 0.20]), the Wald test suggests that MRS differs between the two groups (p=0.102). The point estimates suggest that those in the lowest quartile in terms of VAS scores may place less of a priority upon prioritizing younger patients. MRS for dis2 varies (p=0.0000) between those with lower quartile VAS scores and those in the inter-quartile range (MRS = 10.67 [CI: 7.80 / 13.54] vs. 16.30 [13.54 / 19.07]). Those in the lowest quartile value prioritizing those with moderate rather than severe diseases affecting life expectancy by less than other respondents, which is compatible with self-interested behaviour.

For ill2 (MRS = 3.32 [CI: 1.37 / 5.26] vs. 4.54 [CI: 3.03 / 6.04]) the Wald test (p=0.0049) suggests a difference between the groups, and those with lower quality of life appear to prioritize those with moderate rather than severe diseases affecting quality of life by less. This is again compatible with self-interested behaviour.

Comparing those in the highest quartile in terms of quality of life with those in the inter quartile range, we have already noted a difference in attribute impact and MRS for dep, but MRS might vary for other attributes. For tiss (MRS = 1.24 [CI: 0.50 / 1.97] vs. 1.62 [CI: 1.13 / 2.11]), there is evidence that those with higher quartile VAS scores, place less emphasis upon prioritising those who are a good tissue match, compared to other respondents, according to the Wald test (p=0.0001)

In relation to dis2 (MRS = 18.74 [CI: 13.98 / 23.51] vs. 16.30 [CI: 13.54 / 19.07]) the Wald test (p=0.0000) suggest that MRS varies between those in the highest quartile of VAS scores, and

those in the inter-quartile range. Those in the highest VAS score quartile seem to value prioritizing those with moderate rather than severe diseases affecting life expectancy by more. Also the Wald test (p=0.0069) suggests that MRS differs (MRS = 4.33 [CI: 1.94 / 6.72] vs. 4.54 [CI: 3.03 / 6.04]) in relation to ill2 between those in the highest VAS score quartile and those with a VAS score in the inter-quartile range. Those in the highest quartile value prioritizing those with moderate rather than severe diseases affecting quality of life less. There is no evidence of statistically significant differences in the valuation of any other attributes in terms of MRS.

4.8. Analysis of data for (for respondents who do not claim to only consider others). Comparing the preferences of those with preferences in 1st and 4th quartiles of VAS scores, with the inter-quartile range.

We can consider whether it is appropriate to model preferences separately for altruistic vs. non-altruistic respondents for VAS data, or whether a pooled model is adequate by comparing the restricted pooled model (model 5) with the unrestricted separate models for non-altruistic (model 6) and altruistic (model 7) respondents using a likelihood ratio test. The value of λ (45.31) just exceeds the critical value (40.11) at the 5% level, which suggests that using 2 separate models is also more appropriate for the VAS data.

Table 10 – model 6: <u>For respondents who do not claim to only consider others</u>: dummy variables for 1st & 4th VAS quartiles.

Attribute	Coefficient for inter-quartile range	MRS for inter-quartile range	Coefficient on dummy for those in quartile 1	Implied MRS for lowest quartile	Wald test p-values
wait	.0482**	1	.0097	1	
tiss	.0643**	1.33** (0.64 / 2.03)	-.0299	0.59 (-0.09 / 1.28)	p=0.0144
dep	.0901**	1.87** (1.28 / 2.45)	-.0038	1.49** (0.89 / 2.09)	p=0.0027
age	.0055**	0.11** (0.04 /0.19)	.0039	0.16** (0.08 / 0.24)	p=0.8249
dis1	.0044	0.09 (-2.06 / 2.24)	.1520	2.70* (0.31 / 5.09)	p=0.2437
dis2	.5858**	12.15 (8.65 / 15.65)	-.0555	9.15** (5.77 / 12.52)	p=0.0001
ill1	-.1209	-2.51 (-5.19 / 0.18)	.0060	-1.98 (-4.89 / 0.93)	p=0.4436
ill2	.1537**	3.19** (0.96 / 5.41)	.1615	5.44 (3.02 / 7.86)	p=0.9077
Intercepts	.1422**		.0212		
Attribute	**Coefficient for inter-quartile range**	**MRS for inter-quartile range**	**Coefficient on dummy for those in quartile 4**	**Implied MRS for highest quartile**	**Wald test p-values**
wait	.0482**	1	-.0127	1	
tiss	.0643**	1.33** (0.64 / 2.03)	-.0075	1.60* (0.11 / 3.10)	p=0.0703
dep	.0901**	1.87** (1.28 / 2.45)	-.0341	1.58** (0.42 / 2.74)	p=0.0002
age	.0055**	0.11** (0.04 /0.19)	.0064*	0.33** (0.14 / 0.53)	p=0.8861
dis1	.0044	0.09 (-2.06 / 2.24)	-.0366	-0.91 (-5.39 / 3.58)	p=0.7511
dis2	.5858**	12.15 (8.65 / 15.65)	.1967	22.05** (11.73 / 32.38)	p=0.0273
ill1	-.1209	-2.51 (-5.19 / 0.18)	.0889	-0.90 (-6.58 / 4.78)	p=0.2214
ill2	.1537**	3.19** (0.96 / 5.41)	.0628	6.10* (1.42 / 10.79)	p=0.5509
Intercepts	.1422**		-.0159		
% of actual values predicted:	65.90%	**Sample:**	357 patients: 105 in first quartile and 75 in 4th quartile.	**McFaddens R^2:**	0.119
LR test (λ):	24.09	**Jointly significant?**	No : CV for 18 dfs = 28.87	**Log-likelihood:**	-1942.08

Model 6 (table 10) relates to respondents who do not claim to only consider others (357 / 895 responses). The likelihood ratio test ($\lambda = 24.09$, and the critical value = 28.87 for 18 degrees of freedom) suggests that the dummy variables are not jointly significant. The model accurately predicts 65.90% of responses and McFaddens $R^2 = 0.119$.
Moreover, only 1 / 16 of the interaction dummy variables is significant at the 5% level (age) which relates to a comparison between those in the highest quartile in terms of VAS score, and the inter-quartile range but MRS is not significantly different for this variable.

Comparing those in the lowest quartile with those in the inter-quartile range, MRS varies in a statistically significant way (at the 5% level) for 3 variables according to the Wald test. It varies with respect to tiss (p=0.0144), and MRS (MRS = 0.59 [CI: -0.09 / 1.28] vs. 1.33 [CI: 0.64 / 2.03]) is insignificant for non-altruistic respondents in the lowest VAS quartile. However, it is positive and significant for those in the inter-quartile range (who do value prioritizing those with better tissue matches). It also varies with respect to dep (p=0.0027), and differences in MRS (MRS = 1.49 [CI: 0.89 / 2.09] vs. 1.87 [CI: 1.28 / 2.45]), suggest that non-altruistic respondents in the lowest VAS quartile, would prioritize those with dependents by less than those in the VAS inter-quartile range. There is also a difference with respect to dis2 (p=0.0001), and those in the lowest VAS quartile have a lower MRS to those in the inter-quartile range (MRS = 9.15 [CI: 5.77 / 12.52] vs. 12.15 [CI: 8.65 / 15.65]), meaning non-altruistic respondents in the lowest VAS quartile would not prioritize those with moderate diseases rather than severe diseases affecting quality of life as much as those in the inter-quartile range (with a better quality of life). This finding might be explained by pursuit of self-interest.

Comparing those in the highest quartile with those in the inter-quartile range, MRS varies significantly for 2 variables. MRS varies with respect to dep (p=0.0002) and differences in MRS (MRS = 1.58 [CI: 0.42 / 2.74] vs. 1.87 [CI: 1.28 / 2.45]), suggest that those in the highest VAS quartile have a lower MRS (MRS = 1.58 [CI: 0.42 / 2.74] vs. 1.87 [CI: 1.28 / 2.45]), and would prioritize those with dependents less than those in the inter-quartile range. MRS also varies with respect to dis2 (p=0.0273), and MRS is considerably higher for those in the highest VAS score quartile (MRS = 22.05 [CI: 11.73 / 32.38] vs. 12.15 [CI: 8.65 / 15.65]). Non-altruistic responders who are in the highest VAS score quartile, place more emphasis upon prioritizing those with moderate rather than severe diseases affecting life expectancy, than those in the VAS score inter-quartile range, which could be explained in terms of self interested behaviour. MRS does not vary with respect to any other variables.

4.9. Analysis of data for (for respondents who claim to only consider others, who we label 'altruistic'). Comparing the preferences of those with preferences in 1st and 4th quartiles of VAS scores, with the inter-quartile range.

Model 7 (table 11) relates to altruistic responders, its dummy variables are jointly significant (λ =33.24, whereas the critical value for 18 degrees of freedom is 28.87). The model accurately predicts 64% of responses and McFaddens R^2 = 0.115.Also 3 dummy variables are significant. They include the dummy variables dis2, and ill2 for the lowest VAS quartile group, and dep for the highest VAS quartile group.

Table 11 – model 7: <u>Altruistic patients</u>: dummy variables for 1st & 4th VAS quartiles.

Attribute	Coefficient for inter-quartile range	MRS for inter-quartile range	Coefficient on dummy for those in quartile 1	MRS for lowest quartile	Wald test p-values
wait	.0430**	1	-.0026	1	
tiss	.0789**	1.84** (1.15 / 2.52)	-.0087	1.74** (0.72 / 2.77)	p=0.0077
dep	.0639**	1.49** (0.98 / 1.99)	-.0116	1.30** (0.55 / 2.04)	p=0.0039
age	.0074**	0.17** (0.10 / 0.24)	-.0039	0.09 (-0.01 / 0.19)	p=0.0015
dis1	-.0478	-1.11 (-3.11 / 0.88)	-.0196	-1.67 (-4.74 / 1.40)	p=0.8031
dis2	.8312**	19.34** (15.17 / 23.51)	-.3248**	12.55** (7.52 / 17.57)	p=0.0000
ill1	-.1429*	-3.32** (-5.79 / -0.86)	.0672	-1.88 (-5.69 / 1.94)	p=0.1285
ill2	.2362**	5.50** (3.44 / 7.55)	-.2047*	0.78 (-2.53 / 4.09)	p=0.0001
Intercepts	0.147**		-.0451		
Attribute	**Coefficient for inter-quartile range**	**MRS for inter-quartile range**	**Coefficient on dummy for those in quartile 4**	**MRS for highest quartile**	**Wald test p-values**
wait	.0430**	1	-.0017	1	
tiss	.0789**	1.84** (1.15 / 2.52)	-.0343	1.08* (0.24 / 1.92)	p=0.0005
dep	.0639**	1.49** (0.98 / 1.99)	-.0375*	0.64* (0.00 / 1.28)	p=0.0001
age	.0074**	0.17** (0.10 / 0.24)	-.0009	0.16** (0.06 / 0.25)	p=0.0165
dis1	-.0478	-1.11 (-3.11 / 0.88)	.0760	0.68 (-1.97 / 3.33)	p=0.2368
dis2	.8312**	19.34** (15.17 / 23.51)	-.1127	17.40** (12.10 / 22.70)	p=0.0000
ill1	-.1429*	-3.32** (-5.79 / -0.86)	-.0217	-3.99* (-7.32 / -0.65)	p=0.3804
ill2	.2362**	5.50** (3.44 / 7.55)	-.0882	3.58* (-0.79 / 6.38)	p=0.0034
Intercepts	0.147**		-.0650		
% of actual values predicted:	64.00%	**Sample:**	538 patients: 119 in 1st quartile & 149 in 4th.	**McFaddens R^2:**	0.115
LR test (λ):	33.24	**Jointly significant?**	Yes : CV for 18 dfs = 28.87	**Log-likelihood:**	-.2952.1

Comparing the lowest VAS score quartile with the inter-quartile range, MRS differs with respect to 5 variables for altruistic responders (tiss, dep, age, dis2, and ill2). For tiss (p=0.0077), the, small difference in MRS (MRS = 1.74 [CI: 0.72 / 2.77] vs. 1.84 [CI: 1.15 / 2.52]) suggests that those in the lowest VAS quartile would prioritize potential recipients with better tissue matches marginally less than those with inter-quartile range VAS scores. With respect to dep there is again a significant difference in MRS (p=0.0039), and the difference (MRS = 1.30 [CI: 0.55 / 2.04] vs. 1.49 [CI: 0.98 / 1.99]) implies that those with lower VAS scores have a lower preference for prioritizing those with dependents than those with inter-quartile range VAS scores. For age, MRS again varies significantly (p=0.0015), and MRS (MRS = 0.09 [CI: -0.01 / 0.19] vs. 0.17 [CI: 0.10 / 0.24]) is insignificant for those in the lowest VAS quartile, but positive and significant for those with inter-quartile range VAS scores. There is a difference in MRS (p=0.0000) with respect to dis2, and MRS is lower for those in the lowest VAS quartile relative to the inter-quartile range (MRS = 12.55 [CI: 7.52 / 17.57] vs. 19.34 [CI: 15.17 / 23.51]). This suggests that even amongst altruistic respondents those with poor quality of life (in the lowest VAS quartile), have less of a preference for prioritizing those with moderate rather than severe diseases affecting quality of life, this is a finding that would be easier to explain if the respondents had been non-altruistic.

Also MRS varies for ill2 (p=0.0001) and MRS (MRS = 0.78 [CI: -2.53 / 4.09] vs. 5.50 [CI: 3.44 / 7.55]), it is insignificant for those in the lowest VAS quartile, but positive and significant for those in the inter-quartile range. Again, this is a finding that is hard to explain in a sample of patients that report having considered only others (unless of course they empathise more with the plight of those similarly adversely affected in terms of quality of life to themselves).

Comparing those in the highest VAS quartile with those in the VAS inter-quartile range, MRS varies with respect to 5 variables. For tiss MRS differs significantly (p=0.0005), and MRS

(MRS = 1.08 [CI: 0.24 / 1.92] vs. 1.84 [CI: 1.15 / 2.52]) is lower for those in the highest VAS quartile, so they value prioritizing potential recipients based upon tissue match less. For difdep MRS differs statistically (p=0.0001), and MRS is lower amongst those in the highest quartile of VAS scores compared with the inter-quartile range (MRS = 0.64 [CI: 0.00 / 1.28] vs. 1.49 [CI: 0.98 / 1.99]), so those with high quartile VAS scores value prioritizing those with dependents by less.

With respect to age, the Wald test (p=0.0165) suggests MRS differs. However MRS (MRS = 0.16 [CI: 0.06 / 0.25] vs. 0.17 [CI: 0.10 / 0.24]) is only marginally lower for those in the highest VAS quartile. Also MRS for dis2 differs (p=0.0000), and MRS (MRS = 17.40 [CI: 12.10 / 22.70] vs. 19.34 [CI: 15.17 / 23.51]) is lower amongst respondents who are in the highest VAS quartile, who prioritize those with moderate rather than severe diseases affecting life expectancy by less, which is compatible with altruistic behaviour. Finally, there is evidence that MRS differs with respect to ill2 (p=0.0034) and that MRS (MRS = 3.58 [CI: -0.79 / 6.38] vs. 5.50 [CI: 3.44 / 7.55]) is lower for those in the highest VAS quartile. So those in the highest VAS quartile would prioritize those with moderate rather than severe diseases affecting quality of life less, which is again compatible with altruistic behaviour. There is no evidence that MRS for other variables does differs for this model.

5. Discussion.

Before we discuss this data, probably the most immediate question we need to address is the value of including the Visual Analogue Scale (VAS). There is evidence that VAS tends to elicit different results to those based upon Eq-5d (Robinson, Dolan et al. 1997). Therefore, because it is likely that the links between VAS scores and respondent choices might differ, we chose to use both Eq-5d and VAS. However, the use of VAS has been challenged on a variety of grounds. Firstly, it does not involve a widely accepted measure of utility, and it has therefore been argued it lacks theoretical foundations. Moreover, some attempts to redress this problem by converting VAS scores into standard gamble utilities have foundered (Robinson, Loomes et al. 2001). The claim that VAS lacks theoretical foundations is of course open to question, as it is consistent with the non-welfarist foundations of both QALYs and cost-utility analysis (Parkin and Devlin 2006). Also a case can be made that VAS could play a continuing, limited but useful, role with respect to preference evaluation (Torrance, Feeny et al. 2001; Brazier and McCabe 2007) alongside collection of ordinal data, which is why we included it. It is therefore worth establishing how respondent quality of life as measured by VAS appears to affect preferences, as well as looking at the links between respondent Eq-5d status and respondent preferences.

5.1. Evidence of differences in preferences between the pooled model and separate models according to how respondents replied to the question relating to the perspective they adopted.

The findings of the 2 likelihood tests comparing the pooled models (model 2 and model 5) with the equivalent separate models (models 3 and 4, and models 6 and 7 respectively), suggests that preferences do differ between respondents who do not claim to only consider

others and altruistic responders, both when looking at the links between Eq-5d status and preferences, and VAS status and preferences.

This supports the case for modelling preferences separately for respondents who do not claim to only consider others and altruistic responders. However, for analyses relating to Eq-5d, the within model likelihood ratio tests which test for the joint significance of the dummy variables are insignificant in all 3 models (models 2, 3, and 4). Moreover, attribute impacts do not vary significantly according to which Eq-5d status grouping respondents are in (i.e. bottom quartile, mid-range 47%, or those for whom QALY = 1). The evidence for this is that the dummy variables are never jointly significant in either model 2, 3, or 4, nor do any of them have significant dummy variables.

5.2. Links between respondents Eq-5d status, the perspective they adopted, and their preferences for different transplant prioritization criteria.

We considered that it was conceivable that the relationship between Eq-5d status and preferences for the pooled model (model 2) might have been muted. This is because the majority of the pooled sample of respondents (which includes respondents who do not claim to only consider others and altruistic responders) indicated they had an altruistic perspective. This perspective might blunt the relationship between patient respondent quality of life status (as measured by Eq-5d) and patient preference (particularly for attributes relating to diseases affecting life expectancy, and diseases affecting quality of life). If patients are altruistic it should mean that their own quality of life measured using Eq-5d does not influence their preferences for attributes such as diseases affecting life expectancy or quality of life, based upon their own self interest. This means that associations between quality of life and preferences, which we might expect, might be hidden under an altruistic veil.

Therefore we conjectured that the relationship between respondent Eq-5d status and preferences should be more apparent amongst respondents who do not claim to only consider others (model 3), because low or high Eq-5d status respondents have an incentive to prioritize people for transplant who have similar quality of life characteristics. But, just like the whole sample (model 2), the dummy variables in model 3 proved to be jointly insignificant; and none of the dummy variables were singularly significant in the group that do not claim to only consider others (who are at least partly self interested). This same picture of the dummy variables not being either jointly or singularly coefficient emerges for model 4. There is however some evidence that the MRS figures vary according to which Eq-5d grouping respondents are in. Comparing model 2 (pooled sample), with model 3 (respondents who do not claim to only consider others), 3 / 7 of the MRS figures differ between those in the lowest quartile of MRS in model 2, and the same 3 / 7 are different in model 2. What is interesting though is that, whilst those in the lowest Eq-5d sample would prioritize tiss more in model 2 (whole sample), the respondents who do not claim to only consider others (at least partly self interested) for model 3 in the lowest Eq-5d quartile prioritize those with better tissue matches less.

Comparing model 2 (pooled sample) with model 3 (respondents who do not claim to only consider others) when comparing those in the highest Eq-5d quartile with the 'mid-range' 4 / 7 estimates of MRS are different in model 2 (pooled sample), but only 2 / 7 estimates of MRS differ comparing the top Eq-5d grouping with the mid-range (model 3). The 2 / 7 estimates of MRS which do differ in model 3 also differ in model 2, and the difference in MRS between respondents in the top Eq-5d grouping and the 'mid-range' is in the same direction. At the same time, those in the top Eq-5d grouping, who are in the sample of respondents who do not claim to only consider others (model 3), do not have a statistically significant different

MRS to those in the mid-range for dep and age when there is a difference in model 2 (the whole sample).

Comparing model 2 (the pooled sample), with model 4 (altruistic responders), 3 / 7 MRS figures differ significantly between those in the lowest Eq-5d status quartile and those in the mid-range in the pooled sample. Amongst altruistic responders 2 / 7 MRS figures differ significantly (age and dis2), both of which are significantly different within the pooled sample. Moreover the direction of difference is the same for age, with those in the lowest Eq-5d quartile valuing difage less than those in the 'mid-range.' However, in relation to dis2 those in the lowest Eq-5d quartile who are altruistic have a significantly different MRS compared to those in the 'mid-range.' (MRS =20.43 [12.34 / 28.52] vs. 16.52 [12.85 / 20.18]). This suggests that altruistic responders in the lowest Eq-5d quartile would prioritize those with moderate rather than severe diseases affecting life expectancy by more than those with 'mid-range' Eq-5d status. This cannot be self-interested behaviour but is compatible with people behaving altruistically.

If you take the sample as a whole however (model 2), you find dis2 is valued differently (MRS =14.91 [10.72 / 19.09] vs. 15.03 [12.44 / 17.62]), and respondents in the lowest quartile in the sample overall value prioritizing those with moderate rather than severe diseases affecting life expectancy by less than those in the 'mid-range.' This is compatible with self-interested behaviour by many respondents within the pooled sample.

5.3. Links between respondents VAS status, the perspective they adopted, and their preferences for different transplant prioritization criteria.

In relation to the models which relate VAS status to choice experiment preferences (models 5, 6, and 7), we find evidence that the dummy variables are jointly significant for model 5 (the pooled sample). This provides some evidence that quality of life as measured by VAS has an impact upon practices. Moreover, it appears that 2 / 16 dummy variables are significant, for this model. The likelihood ratio test indicates that the dummy variables are however not jointly significant in model 6 (the non-altruistic responders), and only 1 / 16 of the dummy variables prove significant (age comparing those in the highest quartile with those in the inter-quartile range), which suggests that those in the highest VAS quartile value prioritizing younger respondents more. These findings are surprising because we would expect stronger links between quality of life (as measured by VAS) and patient preferences amongst patients who are non-altruistic (because self interest might encourage such patients to allow their own quality of life to inform their preferences). Equally, it is difficult to understand why the likelihood ratio test for joint significance of the dummy variables is significant for the altruistic group (model 7) when it is not, as we have discussed, for non-altruistic respondents (model 6). Moreover, in model 7 which relates to the altruistic group, 3 / 16 dummy variables are significant, including dis2 when those in the lowest quartile are compared with the inter-quartile range; ill2 when comparing those in the lowest quartile with the inter-quartile range; and dep comparing those in the highest quartile with those in the inter-quartile range.

We found that for non-altruistic responders (model 6), the likelihood ratio test proved the dummy variables to be jointly significant, and 1 / 16 variables proved significant also. So we have the unexpected result that the relationship between VAS status and preferences appears to be stronger for altruistic respondents than non-altruistic respondents.

Finally, in the pooled model (model 5) differences in MRS are statistically significant in relation to 9 / 16 variables. In the non-altruistic group (model 6), 5 / 16 variables are statistically significantly different, and all of these are also significant for the pooled sample (model 5). Overall 3 / 16 of the 5 / 16 variables that have a statistically different MRS relate to differences between the lowest quartile and the inter-quartile range, and these differences are all in the same direction. The other 2 / 16 compare those in the highest VAS quartile with the inter-quartile range, and once again all the differences in MRS between the highest quartile and the inter quartile range are in the same direction.

Comparing the pooled model (model 5) with the altruistic group (model 7), there is evidence that dummy variables are significantly different for 9 / 16 variables in model 5, and for 10 / 16 variables in model 7. All the attributes that are statistically different in the pooled model (model 5) differ amongst altruistic responders (model 7). However, additionally MRS for difage varies amongst the altruistic responders, such that MRS for those in the highest quartile in model 7 for the altruistic group is statistically significantly lower for those in the highest quartile, but only marginally so (MRS =0.16 [0.06 / 0.25] vs. 0.17 [0.10 / 0.24]) for age. The direction of statistically significant differences in MRS is always the same comparing the dummy variable groups with the inter-quartile range except in the case of dis2. For the altruistic sample (model 7) those in the highest quartile of the VAS distribution have a lower MRS for dis2 (MRS =17.40 [12.10 / 22.70] vs. 19.34 [15.17 / 23.51]), meaning those who have better quality of life (in the highest quartile of VAS scores) would place less emphasis upon prioritizing those with moderate rather than severe diseases affecting life expectancy in the altruistic sample, which is compatible with them being altruistic. In contrast in the pooled sample (model 5), MRS is statistically significantly higher (MRS =18.74 [13.98 / 23.51] vs. 16.30 [13.54 / 19.07]), amongst those with better quality of life (in the highest quartile for VAS

scores) for dis2. This is compatible with some self-interested behaviour within the pooled sample (model 5), because those with higher quality of life in this sample are placing more emphasis upon prioritizing people with moderate rather than severe diseases affecting quality of life.

6. Conclusions.

The picture presented above is quite complex. At a methodological level, the main limitation of the analysis presented here is that, ideally, I would have liked to have made a comparison in relation to patient preferences for DCE attributes between respondents who answered in each of the following 3 ways:

1) Answering the questions in terms of what would be best for me.
2) Answering the questions in terms of what would be best for me and others.
3) Disregarding what is best for me and only considering what is best for others.

However, regrettably there were insufficient numbers of respondents (n= 30) in group 1 to enable me to conduct an analysis (using interaction dummy variables) to establish whether their preferences for DCE attributes varied from those in the other two groups. This meant that, when conducting analyses for model 1 (table 5), the 542 respondents who had chosen the third option which could be labelled ‘altruistic’ were compared against patients who had not responded in this way.

Consideration of model 2 (table 6), model 3 (table 7), and model 4 (table 8) leads to some interesting findings. Model 2 was a pooled model in which all respondents providing Eq-5d status information and answering the question about the perspective they adopted were

included (n = 892). Model 3 comprised 359 / 892 respondents who did not claim to only consider others (i.e. the first and second response options), and model 4 comprised the 533 / 892 respondents who had ticked the third response option indicating they only considered others (labelled as 'altruistic'). Since only 30 respondents had indicated that they had responded in an entirely self-interested manner (first response), most of the group of responders who we have described as respondents who 'did not claim to only consider others', in actual fact had an altruistic dimension to their utility function because most of them had considered "what would be best for me and others." Indeed such behaviour is equivalent to the behaviour of an individual who has a 'caring externality.'

The implication of this is that I was unable to compare the preferences of altruistic vs. non-altruistic responders, but rather instead I had to compare the preferences of a group of respondents who claimed to have purely altruistic motives (which I labelled altruistic) with respondents who in contrast considered self interest (although most of this group claimed to consider others at the same time as themselves). Had we achieved a larger sample of respondents who reported they only considered their self-interest then the preferences of purely self interested vs. purely altruistic respondents could have been compared also. Links between the perspective respondents adopted, and how their preferences were affected by their quality of life status, may have been more apparent, if I could have made this more useful comparison.

Exactly the same point can be made in relation to the comparisons between models 5 (table 9), model 6 (table 10), and model 7 (table 11). Model 5 was once again a pooled model, in this case all respondents providing VAS status information as well as answering the question about the perspective they adopted (n= 895). Model 6 comprised 357 / 892 respondents who had not claimed they only consider others (i.e. the first and second response options), and

model 7 comprised the 538 / 892 respondents who had ticked the third response option indicating they only considered others (labelled ‘altruistic’). So once again, because we had not enough respondents who only considered their self interest, I ended up making a comparison between those who claimed to consider only others (labelled altruistic) vs. another group of respondents who took self interest into account; but most of the latter group said they considered others as well as themselves.

Despite the fact that the methodological approach we adopted was compromised because of these sampling issues, there is a case for modelling preferences separately for altruistic vs. other responders. The evidence presented here suggests that patient respondent quality of life as measured by either VAS or Eq-5d influences respondent preferences. However, the links appear to be surprisingly quite weak. When using dummy variables to establish the links between Eq-5d and preferences for model 2 (pooled model), model 3 (those who do not claim to consider only others), and model 4 (altruistic responders), none of the dummy variables proved significant either jointly or individually. However, in terms of the weaker test for differences (statistically significant differences in MRS for variables) there was evidence of some differences in preferences, and some of these differences in the altruistic sample appeared to be compatible with altruistic preferences (model 4). Also, differences in preferences in the non-altruistic sample were compatible with respondents behaving non-altruistically (model 3).

Interestingly, the evidence on links between VAS and preferences is less clear cut. There is some evidence of statistically significant differences in variable impacts in models 5, 6, and 7. However, although some of the differences in MRS seem to be in line with respondents’ non-altruistic or altruistic status, others are not. It could be that part of the reason for this is that those who take into account the interests of others (not themselves), might still have greater

empathy for others who are like themselves. This may explain why the links between quality of life and preferences can still apparently be related to respondents' own quality of life status amongst non-altruistic patients.

It is obviously regrettable from a methodological point of view that we did not have a large enough sample of respondents who claimed to be purely self-interested. Had more respondents indicated they answered the questions in terms of what would be best for themselves we could have established their DCE preferences, and then compared these preferences with the other two groups.

It is not altogether unsurprising that only a small number (n=30) respondents indicated they had answered the DCE questionnaire in terms only of what would be best for them. This is because most of the choices they faced for hypothetical kidney transplant recipients (Patient A or Patient B) would involve patients whose characteristics differed from their own. In such a situation, it is difficult to be purely self-interested when asked to choose which patient should be given a transplant. However, in some other contexts (in which personal utility maximising behaviour might be expected) it should be possible to derive information from enough DCE respondents who take self interest into account.

Despite the limitations arising because we did not have enough purely self interested respondents to assess their group's preferences, there does seem to be some evidence that a patient's own quality of life (as measured by Eq-5d) may influence their DCE preferences, and that less altruistic respondents may be more likely to prioritize respondents with similar characteristics to themselves than altruistic respondents do. However, when undertaking a similar analysis using respondents' VAS scores we found that sometimes the results were counter-intuitive, and not always supportive of the view that altruistic patients are less inclined

to prioritize potential recipients with similar characteristics to themselves, than those with more self-interested motivations.

In conclusion, I think that the hypothesis we are trying to test ideally needs to be tested using a DCE which can obtain sufficient respondents who fall into all 3 categories of behaviour. I am currently conducting a DCE study (in collaboration with Dr Verity Watson, University of Aberdeen) in which I am examining preferences for different modes of renal dialysis. For the purposes of this labelled choice DCE, patients have again completed Eq-5d and VAS, and have been asked a similar question to ascertain whether self-interested or altruistic preferences (or both types of preferences simultaneously) apply. Because we have enough respondents falling into each of the 3 categories, I will be able to analyse the preferences from respondents in the three groups. This should therefore provide a very promising environment in which to examine the issue discussed above in greater detail i.e. the links between respondent quality of life, respondent DCE preferences, and how respondent DCE preferences might be affected by how altruistic a respondent is.

The issue of the extent to which DCE respondents are motivated by self-interest or altruism, and how this might affect respondent preferences, has been given little attention to date in the DCE literature. The analysis presented in this chapter represents the first serious attempt to address this issue. As indicated, I am also now following up this work in a further, separate DCE study. I am confident that publishable findings should emerge from this work.

7. References.

Aspinall, P. A., Z. K. Johnson, et al. (2008). "Evaluation of quality of life and priorities of patients with glaucoma." Invest Ophthalmol Vis Sci **49**(5): 1907-1915.

Brazier, J. and C. McCabe (2007). "'Is there a case for using visual analogue scale valuations in CUA' by Parkin and Devlin - A response: 'Yes there is a case, but what does it add to ordinal data?'." Health Econ **16**(6): 645-647.

Bryan, S., G. Weatherburn, et al. (2001). "The cost-effectiveness of magnetic resonance imaging for investigation of the knee joint." Health Technol Assess **5**(27): 1-95.

Byrne, C., D. Ford, et al. (2008). "ESRD incident rates in 2008: national and centre-specific analyses." **Chapter 3, UK Renal Registry Report [http://www.renalreg.org]**.

Byrne, C., R. Steenkamp, et al. (2008). "ESRD prevalent rates in 2008 national and centre-specific analyses." **Chapter 4, UK Renal Registry report [http://www.renalreg.org]**.

Culyer, A. J. (1976). Need and the National Health Service : economics and social choice. Totowa, N.J., Rowman and Littlefield.

Culyer, A. J. (1980). The political economy of social policy. New York, St. Martin's Press.

Dowie, J. (1985). "The political economy of the NHS: individualist justifications of collective action." Soc Sci Med **20**(10): 1041-1048.

Gill, P. and L. Lowes (2008). "Gift exchange and organ donation: donor and recipient experiences of live related kidney transplantation." Int J Nurs Stud **45**(11): 1607-1617.

Grant, A., S. Wileman, et al. (2008). "The effectiveness and cost-effectiveness of minimal access surgery amongst people with gastro-oesophageal reflux disease - a UK collaborative study. The REFLUX trial." Health Technol Assess **12**(31): 1-181, iii-iv.

Grossmann, V. (2002). "Is it rational to internalize the personal norm that one should reciprocate?" Journal of Economic Psychology **23**(1): 27-48.

Haltiwanger, J. and M. Waldman (1993). "The Role of Altruism in Economic Interaction." Journal of Economic Behavior & Organization **21**(1): 1-15.

Hjelmgren, J. and A. Anell (2007). "Population preferences and choice of primary care models: A discrete choice experiment in Sweden." Health Policy **83**(2-3): 314-322.

Hudson, J. and P. R. Jones (1994). "The importance of the 'ethical voter'. An estimate of 'altruism'." European Journal of Political Economy **10**: 499-509.

Jacobsson, F., J. Carstensen, et al. (2005). "Caring externalities in health economic evaluation: how are they related to severity of illness?" Health Policy **73**(2): 172-182.

Johansson, P. O. (1994). "Altruism and the value of statistical life: empirical implications." J Health Econ **13**(1): 111-118.

Khalil, E. L. (2004). "What is altruism? A reply to critics." Journal of Economic Psychology **25**(1): 141-143.

Ossa, D. F., A. Briggs, et al. (2007). "Recombinant erythropoietin for chemotherapy-related anaemia: economic value and health-related quality-of-life assessment using direct utility elicitation and discrete choice experiment methods." Pharmacoeconomics **25**(3): 223-237.

Parkin, D. and N. Devlin (2006). "Is there a case for using visual analogue scale valuations in cost-utility analysis?" Health Econ **15**(7): 653-664.

Patel, S. R., P. Chadha, et al. (2011). "Expanding the live kidney donor pool: ethical considerations regarding altruistic donors, paired and pooled programs." Exp Clin Transplant **9**(3): 181-186.

Ratcliffe, J., M. Buxton, et al. (2005). "Determining priority for liver transplantation: a comparison of cost per QALY and discrete choice experiment-generated public preferences." Appl Health Econ Health Policy **4**(4): 249-255.

Robinson, A., P. Dolan, et al. (1997). "Valuing health status using VAS and TTO: what lies behind the numbers?" Soc Sci Med **45**(8): 1289-1297.

Robinson, A., G. Loomes, et al. (2001). "Visual analog scales, standard gambles, and relative risk aversion." Medical Decision Making **21**(1): 17-27.

Schenk, R. E. (1987). "Altruism as a Source of Self-Interested Behavior." Public Choice **53**(2): 187-192.

Sen, A. K. (1977). "Rational Fools - Critique of Behavioral Foundations of Economic-Theory." Philosophy & Public Affairs **6**(4): 317-344.

Siminoff, L., M. B. Mercer, et al. (2007). "The reasons families donate organs for transplantation: implications for policy and practice." J Trauma **62**(4): 969-978.

Torrance, G. W., D. Feeny, et al. (2001). "Visual analog scales: Do they have a role in the measurement of preferences for health states?" Medical Decision Making **21**(4): 329-334.

Trivers, R. L. (1971). "Evolution of Reciprocal Altruism." Quarterly Review of Biology **46**(1): 35-&.

Wooldridge, J. M. (2002). "Economic analysis of cross section and panel data." MIT press.

Chapter 9: Discussion, and Conclusions.

1. Introduction.

This thesis has presented some of the research I have conducted over a number of years, specifically focused on the use of Discrete Choice Experiments (DCEs). Earlier chapters have described different studies and how my approach towards the use of DCEs has developed, as well as setting this research in the context of the developing literature for use of DCEs in health. In this chapter, I first briefly recap on the previous chapters and the conclusions drawn. This is followed by pulling the themes together and discussion of the key methodological issues emerging. The use of DCE to inform policy making is still in its infancy. At the same time, the popularity of this powerful technique is increasing. Now is a good time to consider its use to date and its future value.

2. A developing evidence base.

In Chapter 1, I provided a definition of DCEs and explained key steps involved in undertaking a DCE (section A); and demonstrated that DCEs have theoretical foundations within the characteristics theory of demand (section B) and Random Utility Theory (section C). I then completed a systematic review of the DCE literature relating to health and presented my findings (section D). This review encompassed evidence from earlier systematic reviews for the periods 1990 – 2000 (Ryan and Gerard 2003) 2001 – 2008 (de Bekker-Grob, Ryan et al. 2012) plus 96 more recent studies which were reviewed and analyzed by me for 2009-2011.

The review began by considering which categories DCE studies fall into, and then provided an overview of trends in the number of studies published per year. This number has increased exponentially from 3 per year during 1990 – 2000, to 14 per year during 2001 – 2008, and then 32 per year during 2009 - 2011. Information was also presented relating to the country of origin for DCE studies. A key finding here was that the UK's previous dominance (48% of DCE studies in 2001 – 2008) as the major source of published DCE studies appears to be threatened. The UK now accounts for just 21% of studies. This is primarily due to increasing numbers of studies emanating from countries such as the USA, Canada, Denmark, the Netherlands, Germany, and other countries. On a positive note, this means that DCE are increasingly being applied internationally, while the actual annual number of DCEs in the UK (55 / 114 from 2001 – 2008; and 20 / 96 from 2009 – 2011) has not changed much in the last decade (e.g. around 8 a year from 2001 – 2008, and 7 a year from 2009 – 2011). Moreover, DCEs have the advantage over Cost-Utility Analysis that they can value both health outcomes and experience factors (and hence value process utility). In the UK, whilst many health economists regard DCEs as a useful tool for economic evaluation, Cost-Utility Analysis still seems to be used more often alongside randomized controlled trials. This partly reflects the fact that policy makers were already familiar with the use of Cost-per-QALYs to help inform decision making, before DCEs had become established. Before this technique is used to underpin more decision making in healthcare, I think more DCE studies using WTP to underpin Cost-Benefit Analysis will need to be undertaken. In this way policy makers might begin to think in terms of what constitutes an acceptable Cost-Benefit ratio, not acceptable just an acceptable Cost per QALY.

However, this thesis points to wide ranging methodological issues surrounding the current use of DCEs to elicit WTP (Chapter 1, 3, and 4). In particular the need to ensure that DCE questionnaires are designed to be incentive compatible. Moreover, the range set for DCE

monetary attributes should be appropriate and informed by suitable pilot DCE research. Finally, DCE attributes and levels need to be carefully selected (a process informed by thorough qualitative research), so that benefits can be calculated for all the important characteristics that patients value.

Trends are also observable in the number of attributes in DCE studies. A key finding here is that during 2009 – 2011 there were fewer studies with low numbers of attributes (i.e. either 2 – 3, or 4 – 5, and an increased number of studies with a large number of attributes (i.e. 6 – 9 attributes). Interestingly, when I examined the domains of DCE attributes, it was evident that the number of studies with a monetary measure has remained reasonably constant, although increasing numbers of studies appear to have attributes relating to risk, health status, and other domains. There also seems to be an increasing trend towards presenting respondents with more DCE questions; in 2009 – 2011 only 23% of studies had 8 or less choices, and increasing numbers of studies (55% had 9 – 16 choices, with 22% presenting more than 16 choices. Some evidence suggests (Bech, Kjaer et al. 2011) that later DCE responses might be more thought through, so this may be a welcome development. There has also been a shift away from self-administered DCE questionnaires towards more interviewer administered ones and, particularly, computer administered questionnaires. The fact that DCEs are increasingly computer administered is inevitable given technological developments. It may also help DCE researchers more easily access respondents. However, care must be taken to ensure that this does not exclude certain potential respondent groups (in the interests of ensuring sample representativeness) just because they are less computer literate or do not have access to a computer e.g. older people.

Obviously, having an appropriate DCE experimental design is extremely important. In section D.4 I outlined a range of designs that can be used. Main effects designs (which do not cater

for interactions) still accounted for the bulk of designs in 2009 – 2011. However, reassuringly, an increasing proportion of studies are now catering for interaction effects (11% in 2009 – 2011, up from 5% in 2001 – 2008).

The use of software packages to design DCEs had fallen to 46% of studies by 2009-2011. Use of Sawtooth and SPSS has not changed very much. However, there has been a shift away from the somewhat dated package (SPEED) which is used to generate orthogonal designs, towards the use of SAS which can be used to generate D-optimal designs. D-optimal designs have the advantage that they are more statistically efficient than many orthogonal designs, which means information on preferences can be inferred from a smaller sample of respondents. That said, orthogonal designs using suitable software packages or design catalogues can still provide competent DCE designs, and may be sufficient if researchers already plan to distribute a large number of questionnaires (as was the case for all the studies reported in this thesis). Design catalogues are increasingly used, and in 2009 – 2011 accounted for 16% of designs. In the same period, 4% of DCEs used websites for DCE designs, and 6% used expert advice. Worryingly, however, around 27% of studies published in 2009-2011 failed to adequately report the source of their design.

In terms of analysis, a variety of different econometric methods can be used to analyse DCE data (see section D.5 of Chapter 1). All the DCEs reported in this thesis involved a choice between just two options (A or B), and so the data from them could be analyzed using binary dependent variable models. I therefore used Random Effects Probit for estimation of most of the models (Chapters 2, 3, 4, 5, 6, and 8). When each of the studies in this thesis was piloted (i.e. before 2007), the use of Random Effects Probit was very common, and remained common until recently. So for example during 1990 – 2000, 53% of studies were Random Effects Probit. However, use of such models then started to decline, falling to 41% in 2001 –

2008, and down further to 15% in 2009 – 2011. By 2009 – 2011, other binary dependent models such as Random effects Logit accounted for 9% of DCEs (in the same period Probit accounted for just 2% of studies and Logit 9%). There has been a corresponding increase in the use of Multinomial Logit models (18% in 1990 – 2000, 22% in 2001 – 2008, up to 45% in 2009 – 2011). Part of the reason for this is that DCEs are increasingly designed to allow respondents to select between more than 2 options, or between two options plus an opt-out. In some situations presenting more than 2 options is essential. An example of this is the new labeled choice DCE I am currently conducting which compares 3 models of dialysis provision (hospital based haemodialysis, home based haemodialysis, and peritoneal dialysis). If I had not allowed for at least 3 response options, I would not have been able to adequately present a sufficient range of dialysis options to respondents.

However, the use of a binary dependent model such as Random Effects Probit was adequate for the purposes of evaluating 2 competing models of DVT provision (Chapter 2 of this thesis); or 2 competing models of gynaecological care provision (Chapters 3 and 4); or for evaluating preferences for renal transplantation (Chapters 5, 6, and 8). A limitation of this approach is that information on preference heterogeneity is not provided (except if interaction dummy variables are used for respondent sub-groups). Of course, I could have given respondents the option not to make a choice. However, in some contexts (for example preferences for renal transplantation) that might not be appropriate. During piloting I discovered that a number of respondents who were not indifferent would have preferred to avoid registering a choice, simply because they felt uncomfortable with ‘playing God’. I therefore decided that in order to obtain as much information about preferences as possible, I ought to force a choice.

The use of Nested Logit might be indicated if there is a need to relax the Independence of Irrelevant Alternatives (IIA) assumption, but even by 2009 – 2011 only 3% of studies used this form of analysis. I think researchers do need to be sure that the IIA assumption holds before simply using Multinomial Logit (MNL). I am aware that for analysis of my new renal dialysis DCE I may have to use Nested Logit because two of my response options are quite similar (i.e. hospital based haemodialysis and home based dialysis) whilst the third option (home based peritoneal dialysis) is rather different, so the IIA option assumption may not hold. Had I not read recent key literature relating to econometric analysis of DCE data (de Bekker-Grob, Ryan et al. 2012) it would have been easy to make the potential mistake of unthinkingly adopting MNL, because it is "what other researchers appear to be doing".

The findings of my systematic review of the literature show that there has been a big rise in the application of Mixed Logit (this accounted for 18% of studies by 2009 – 2011). Mixed Logit is relatively easy to undertake and it allows researchers to establish whether preferences for particular attributes are subject to preference heterogeneity or not. When either Mixed Logit or LCM is applied, my literature review suggests it always seems to provide at least some evidence of preference heterogeneity. It could be argued that, since both Mixed Logit models and LCM invariably seem to indicate that preferences are heterogeneous, there is little point in applying them because they generate a result which can be anticipated (i.e. evidence of preference heterogeneity). I think, however, that such an argument misses two crucial points. The first is that by applying either Mixed Logit or LCM, it is possible to establish which attributes might be valued more differently by respondents, and which parameters might be less susceptible to preference heterogeneity. Secondly, if information about preference heterogeneity is not obtained, any results may give the misguided impression that respondent preferences are similar, when in reality preferences are heterogeneous across respondents. For these reasons I think that researchers need to

consider using either Mixed Logit or LCM routinely to analyze DCE data (perhaps alongside other econometric models).

The use of validity checks when undertaking DCEs has been examined in some detail in chapter 1 of this thesis. The 'Gold standard' for validity checks requires some assessment of external validity. This involves a comparison of stated preferences obtained using DCEs with revealed preference data on actual choices. Only one published DCE paper over the period 1990 – 2011 has been identified using such an approach (Mark and Swait 2004) for a healthcare DCE. So, it is apparent that tests of external validity are extremely rare for healthcare DCEs. Applying tests for DCE external validity in healthcare, therefore, remains an under-researched area. Tests for external validity might be undertaken by applying DCEs in some contexts in which people have to pay for health provision, and comparing the DCE results with revealed preferences. There is an interesting discussion paper (Watson V and Ryan M 2010) which compares the monetary value generated from a DCE model for Chlamydia screening with the real price of a Chlamydia screening test introduced at a chemist. However, since market prices don't necessarily equate with someone's full WTP, this still does not provide a full test of external validity (de Bekker-Grob, Ryan et al. 2012).

In contrast with the extreme rarity of tests of external validity, I found evidence that an increasing proportion of published studies report tests for internal theoretical validity (72% in 2009 – 2011). This type of validity test requires that study authors assess whether coefficients upon attributes move in line with prior expectations. However, at the same time there has been a general decline in the use of a range of other validity tests since 2001-2008. These include tests to establish whether preferences are transitive (9% of studies in 1990 – 2000, 4% in 2001 – 2008, but 0% in 2009 – 2011); non-satiation (44% of studies in 1990 – 2000, 49% of DCEs in 2001 – 2008, down to 15% in 2009 – 2011); compensating decision

making (35% in 1990 – 2000, 32% in 2001 – 2008, down to 15% in 2009 – 2011); and Sen's extraction and contraction properties (0% in 1990, 2% in 2000-2008, and 1% in 2009 – 2011). This decline may be related to the observation (Lancsar and Louviere 2006) that some behaviour that DCE researchers have identified in rationality tests as apparently irrational, may not be out of step with rational decision making. Moreover, some analyses which have used qualitative research techniques (Miguel, Ryan et al. 2005, Ryan, Watson et al. 2009) have found that individuals defined as failing non-satiation and Sen's expansion and contraction properties in quantitative tests had 'rational' reasons for doing so (de Bekker-Grob, Ryan et al. 2012). Furthermore, it has been pointed out (Lancsar and Louviere 2006) that random utility models are robust to both violations of compensatory decision making and errors made by individuals in forming and revealing preferences (de Bekker-Grob, Ryan et al. 2012). This body of evidence suggests that the decline in the use of many validity tests partly reflects the fact that researchers increasingly feel they do not yield useful information.

I was however expecting to find increasing evidence that researchers are making use of qualitative methods to improve DCE validity. This was because key publications since 2007 have highlighted the need to deploy qualitative methods to enhance DCE process and design (Coast and Horrocks 2007, Ryan, Watson et al. 2009, Coast, Al-Janabi et al. 2012, de Bekker-Grob, Ryan et al. 2012). However, worryingly the use of qualitative methods appears to be declining, after an initial rise. In 1990-2000, 18% of studies used qualitative methods to inform attribute selection, rising to 69% in 2001 – 2008 before falling back to 38% in 2009 – 2011. A similar trend is apparent in relation to attribute level selection. In 1990 – 2000, 18% of studies used qualitative methods to inform attribute level selection, rising to 33% in 2001 – 2008 before falling back to 25 / 96 (26%) in 2009 – 2011. This pattern is particularly worrying because unless you obtain information from members of the respondent groups to be targeted with DCE questionnaires about attributes and level appropriateness, there is a

danger that attributes that ought to be valued as part of a DCE might be omitted, leading to omitted variable bias. Also, establishing appropriate levels for attributes is important, and especially so if you are using a monetary attribute to try to establish respondents WTP. Without good qualitative research to inform the selection of attributes and levels included in a DCE, a study can end up framing DCE questions inappropriately, thereby compromising DCE validity.

Similarly, use of pre-testing DCE questionnaires has also fallen over the period (pre-testing questionnaire used in 47% of studies in 1990 – 2000, 32% in 2001 – 2008, and 24% in 2009 – 2011). In all the DCE studies I have undertaken, I have always used a pilot pre-testing questionnaire. Data from the pilot questionnaires can be analyzed econometrically, and if particular attributes or attribute levels are not statistically significant then it is possible to re-think whether they ought to be included in the DCE design. I think it is worrying that under a quarter of DCE studies in the most recent period (2009 – 2011) reported that they used pilot pre-testing questionnaires. There also continues to be very little application of measures to strengthen face validity through the use of debriefing choices (0% in 1990 – 2000, 4% in 2001 – 2008, and 2% in 2009 – 2011).

3. Adding to the evidence base.

Considering the work reported in this thesis, Chapter 2 relates to the very first DCE I conducted about 10 years ago (with assistance from Dr Emma Mc Intosh, University of Oxford). Some of this work has been published (Clark, Moro et al. 2009). Chapter 2 demonstrates how DCEs can be used to calculate WTP. The primary objective was to evaluate whether healthcare services for patients with suspected DVT ought to be provided in the community or a hospital outpatient setting. The DCE valued attributes that differed across

the two modes of provision in terms of WTP. A key finding of this chapter is that how DVT provision performs in terms of certain attributes (i.e. speed of diagnosis) might be more important than the location of service provision (i.e. community versus hospital). This finding holds regardless of whether MWTP, uptake rates, or Compensating Variation are used to inform decision making.

Chapter 2 illustrates a major advantage of DCEs, the possibility of calculating WTP for different models of service provision and using MWTP figures to value different models of provision (if a 'state of the world' model is appropriate). Alternatively, it is possible to establish the popularity of different hypothetical service models using uptake rates; or to calculate welfare changes using the compensating variation formula. Clearly therefore, if DCEs can be used to accurately calculate WTP, they may be very useful. Information on WTP for benefits can also inform the benefits side of a Cost-Benefit Analysis.

Having indicated the usefulness of estimates of WTP obtained using DCEs in Chapter 2, in Chapters 3 and 4 I considered whether estimates of WTP obtained using DCEs when healthcare is free at the point of use are likely to be subject to hypothetical bias. Analyses in these chapters relate to research that originally began in 2003. Since 2003, a number other analyses have been published relating to whether estimates of WTP obtained using DCEs might be subject to hypothetical bias. In Chapters 3 and 4 I reported the findings of two separate DCE analyses which included a question to establish whether or not people take a monetary attribute into account when answering choices. I was concerned that, because my respondents received healthcare free at the point of use, a proportion of them might disregard the monetary attribute when they made DCE choices because they knew that in reality a cost would not apply. If respondents fail to take differences in the levels of the monetary attribute into account then this could bias WTP estimates. I therefore tried to establish what proportion

of respondents might fail to take differences in the monetary attribute into account. My logic was that rational respondents should value money, therefore if their responses were not subject to hypothetical bias they should take differences in the levels of the monetary attribute into account when choosing. Moreover, if this was not happening it might indicate estimates of marginal willingness to pay (MWTP) are subject to hypothetical bias. In two similar analyses, using different DCE data (Chapter 3 and Chapter 4), I obtained the result that around a third of DCE respondents indicated that they had not taken differences in the monetary attribute into account when making DCE choices. In both chapters, I then ran econometric models using interaction dummy variables to establish whether estimated WTP varied between respondents who indicated that they had not taken differences in the monetary attribute into account, versus other respondents. Both analyses found evidence that estimated MWTP varied (statistically significant result using Wald tests) for some attributes in the group that claimed they had not taken differences in the monetary attribute into account. This suggests that having a group of respondents who fail to take differences in the monetary attribute into account may bias estimates of MWTP.

One of the problems with this analysis is that a failure to take a monetary attribute into account may have arisen because the levels I set for the monetary attribute may have been inappropriate. This could well be the case because there is evidence that estimated MWTP may be sensitive to the levels assigned to the monetary attribute (Skjoldborg and Gyrd-Hansen 2003). Therefore, if I were to repeat a similar analysis again I would plan to use mainstream WTP analysis during piloting (perhaps using payment cards) to more robustly determine the levels for the monetary attribute. Also, it would be worth having different versions of the questionnaire with different ranges for the levels for monetary attribute in order to establish whether a failure to take differences in the levels of the monetary attribute into account, is related to the levels assigned to the monetary attribute. Moreover, the way in

which the monetary attribute was framed in the questionnaire was not incentive compatible. The questionnaire pre-amble asked respondents to “Please assume that you would lose this amount even if you would not.” This request was made because, in the DCE pilot exercise, it was clear some respondents ignored the monetary attribute because they considered that since healthcare is free at the point of use, a ‘cost to you’ would not apply. However, this wording may have only served to make matters worse by reminding respondents that in reality such a difference in the monetary attribute would not apply. Thus the analyses in Chapters 3 and 4 may have generated an extreme result. Had I framed the monetary attribute in a more incentive compatible manner, and taken more care to establish an appropriate payment vector, I would have almost certainly discovered that the proportion of respondents failing to take the monetary attribute was lower. That said, I was correct to think that there may be cause for concern that estimates of WTP obtained using DCEs can be biased. An analysis (Skjoldborg and Gyrd-Hansen 2003) had shown that estimates of WTP may be sensitive both to the range specified for the monetary attribute and to the presence or absence of payment *per se*. A more recent analysis (Gyrd-Hansen and Skjoldborg 2008) reached the conclusion that respondents might be more influenced by the presence or absence of a non-zero cost than by the level of cost indicated by the monetary attribute.

There is other evidence to indicate that, unless ‘cheap talk’ is used to ensure respondents consider the levels of a monetary attribute when making DCE choices, DCE responses may be insensitive to changes in the levels of the monetary attribute (Ozdemir, Johnson et al. 2009). So ‘cheap-talk’ may be required to more accurately estimate WTP. There is also evidence to suggest that DCE cost functions may not be linear, something which tends to be assumed when conducting MWTP analysis (Ozdemir, Johnson et al. 2009). However, use of ‘cheap talk’ can help to ensure that the cost function becomes linear (Ozdemir, Johnson et al. 2009). Other evidence suggests that respondents might deploy heuristics to recode costs into

categories such as low, medium, or high (Johnson, Mohamed et al. 2011), which inevitably means that estimates of WTP obtained from respondents who deploy such heuristics are less accurate. Also, preferences may be subject to a 'learning curve', so later responses in a DCE might be a better indicator of preferences than earlier responses (Johnson and Desvousges 1997, Carlsson and Martinsson 2001). This fits with evidence that estimated WTP might be affected by the number of DCE choices that respondents face (Bech, Kjaer et al. 2011). This implies that later responses might provide more accurate indications of WTP than earlier ones. Finally, there is also evidence that framing effects might affect estimated WTP i.e. a question framed in terms of number of polyps found versus number of polyps missed may be associated with differences in the absolute levels of MWTP (Howard and Salkeld 2009).

Given the wide ranging concerns surrounding the robustness of eliciting WTP using DCEs, researchers could decide to use DCEs but avoid using them to elicit WTP, in favour of adopting other summary DCE outcome measures (see Chapter 1, section 7.2). However a disadvantage of this approach is that it means estimates of WTP are not generated to inform the Benefits side of a Cost-Benefit Analysis (CBA). An alternative approach would be to use DCEs to elicit information on WTP but take measures to minimize bias. For example, by defining the monetary attribute in a manner which is as incentive compatible as possible and undertaking rigorous qualitative research using mainstream WTP analysis to establish an appropriate price vector for the monetary attribute. Different price vectors for the monetary attribute can be used within the same DCE to see how estimated WTP is affected. This information could inform sensitivity analysis around WTP values (Ryan and Wordsworth 2000). Also, to ensure that respondents take the monetary attribute into account when making choices, researchers could use 'cheap talk' or graphic representations of attribute cost in questionnaire preambles to improve validity (Johnson, Mohamed et al. 2011). Another possibility is the use of a variant of the question I posed about whether respondents take

differences in the monetary attribute into account, after each DCE choice. This would be to ensure respondents take differences in the monetary attribute into account when making DCE choices. My thinking is that it may not possible to ensure questionnaires are completed in interviews using 'cheap talk'. So, as an alternative, a DCE questionnaire pre-amble could say something to the effect that "It is important when making choices that you consider differences in the levels of price. Therefore we are going to ask you to tick a box after each choice scenario to indicate whether you have done this or not." After each DCE question respondents could be then asked "Did you look to see whether price differed between options when making choices" and they could answer 'Yes' or 'No'. This approach may concentrate respondents mind (like 'cheap talk'), so they consider differences in the monetary attribute when choosing.

Chapter 4 also addressed another WTP methodological issue. During DCE piloting consideration was given to how the monetary attribute ought to be described to respondents. The 3 most highly ranked options were 'Cost to you', 'Amount Lost' and 'Willingness to Pay.' As 'Cost to you' was the highest ranking descriptor we used that descriptor for the 'Period Problem' questionnaires (Appendix C). All the questionnaires used for the DCE analysis contained in Chapter 3 had a 'Cost to you' monetary attribute. However, in Chapter 4 I wanted to establish whether adopting different descriptors for the monetary attribute might affect estimated MWTP. I distributed equal proportions of questionnaires with DCE choices which were otherwise the same except for the fact that the descriptor for the monetary attribute varied. It was either 'Cost to you', 'Amount Lost' or 'Willingness to Pay'. I was concerned that a descriptor which referred to cost (like 'Cost to you') might be more likely to induce cost based responses (Ratcliffe 2000) e.g. respondents would value differences in other attributes in terms of how much they think they might cost, rather than how much they value them. This would be of concern, because cost-based valuations might exclude

‘consumer surplus.’ I reasoned that a descriptor such as ‘Willingness to Pay’ uses wording which makes people think in terms of their maximum valuation of something. Therefore, if the ‘Cost to you’ descriptor encouraged cost based prices, I would expect to find that estimates of MWTP obtained using a ‘Willingness to Pay’ descriptor would be higher than those obtained using a ‘Cost to you.’ In fact, Wald tests suggested that estimated MWTP for all attributes was never statistically significantly different, irrespective of which monetary descriptor was used. This was reassuring and would appear to indicate that, so long as the choice of monetary descriptor does not affect who pays, it can be described in different ways without necessarily affecting estimated MWTP.

A separate series of chapters (Chapters 5 – 8) of the thesis were related to the renal transplant DCE study. In Chapter 5 the preferences of renal patients for kidney allocation criteria were ascertained using a DCE. The econometric model used interaction dummy variables to establish whether the preferences of other stakeholder groups (including healthcare professionals, live kidney donors / relatives of deceased donors, and carers) differed from those of patients. The attribute relating to valuing how long people waited for a transplant was used as the denominator for marginal rates of substitution (MRS). Differences in preferences between stakeholder groups were assessed in two ways. The first involved assessing whether there were statistically significant differences in the value of particular attributes, comparing one of the other stakeholder groups (healthcare professionals, live kidney donors / relatives of deceased donors, or carers) with the patient group using interaction dummy variables. The second way of establishing whether preferences differ was to look for statistically significant differences in MRS for attributes, comparing patients with other stakeholder groups using Wald tests (a method not generally used in DCE research). Key findings were that patients appeared to value a range of transplant allocation criteria in a manner compatible with theoretical validity. Moreover, there was evidence of statistically

significant differences in preferences across different stakeholder groups. Of particular importance was the finding that healthcare professionals had preferences which sometimes differed from those of patients. Some of our findings were consistent with changes to kidney transplant allocation policy made in 2006 which placed more emphasis on prioritizing long waiters and young adults. However, based on the DCE findings prioritizing recipients with dependent children or adults might also be considered when UK transplant policy is next re-evaluated.

Chapter 5 also compared the preferences of ethnic minority patients with non-ethnic minority patients. Ethnic minority patients are likely to be doubly disadvantaged if transplants are allocated according to the closeness of donor recipient tissue match. This is because not only are they more likely to require a transplant but there is also a paucity of matched organs for ethnic minority groups. So another key finding (Chapter 5) was that whilst non-ethnic minority groups would prioritize transplanting to patients with a good donor-recipient tissue match, ethnic minority patients would not. This analysis has been submitted to 'BMC Nephrology', who invited us to re-submit it with minor amendments (to tailor it to a clinical audience). Once published, it will form part of the body of evidence used when UK renal transplant policy is re-appraised.

Chapter 6 examines diversity issues in greater depth. The finding reported in Chapter 5 that preferences differ between 'ethnic minority' patients and 'non-ethnic minority' patients is very important. However, the 'ethnic minority' category used in Chapter 5 was broad and included white ethnic minority patients. At the same time, the ethnic minority groups most disadvantaged if kidneys are allocated to closely matching recipients are non-white ethnic minority groups, especially South Asians, because of the high prevalence of diabetes which causes renal disease, and a shortage of kidney donors from these communities. Chapter 6

sets out these diversity issues in some detail, shows how preferences vary between non-white ethnic minority patients vs. other patients; and also Asian patients vs. other patients. Data analysis demonstrated that preferences differed for a number of attributes between groups. The most important finding was that, like ethnic minority respondents overall (Chapter 5), findings for non-white ethnic minorities vs. Asian ethnic minorities (Chapter 6) also indicated that these ethnic minority patient groups did not value prioritizing kidney transplants on the basis of tissue match. In Chapter 6 the issue of gender differences in preferences was also addressed, finding only very limited evidence that patient preferences vary by gender. The chapter concluded that catering for diversity in terms of gender is not necessary, but that substantive differences in preferences exist between non-white ethnic minority patients and other patients and Asian and non-Asian patients. It is concluded that these will need to be considered when transplant policy is next re-appraised. Much of the analysis contained in this chapter has been published (Clark, Gumber et al. 2009), so it can inform kidney transplant policy when it is next re-appraised.

Following on from this work in Chapter 7 Mixed Logit and Latent Class Models (LCM) were used to analyze the renal transplant patient DCE data. Recently a number of discrete choice experiment analyses have been conducted which also assess preference heterogeneity using either Mixed Logit or Latent Class Models. During 2001 – 2008 there were 6 such studies (Mark and Swait 2004, Hall, Fiebig et al. 2006, Lancsar and Louviere 2006, Goto, Nishimura et al. 2007, Lancsar, Hall et al. 2007, Bellary, O'Hare et al. 2008) which used Mixed Logit (de Bekker-Grob, Ryan et al. 2012). One of these studies (Lancsar and Louviere 2006) also conducted LCM alongside Mixed Logit (the only paper using LCM analysis during 2001 – 2008).

During 2009 – 2011 there have been far more published analyses using Mixed Logit (Eberth, Watson et al. 2009, Hauber, Mohamed et al. 2009, Howard and Salkeld 2009, Ozdemir, Johnson et al. 2009, Regier, Friedman et al. 2009, van Helvoort-Postulart, Dellaert et al. 2009, van Helvoort-Postulart, van der Weijden et al. 2009, Blaauw, Erasmus et al. 2010, de Bekker-Grob, Hofman et al. 2010, Johnson, Ozdemir et al. 2010, Scuffham, Whitty et al. 2010, Wittink, Cary et al. 2010, Goto, Takahashi et al. 2011, Mohamed, Epstein et al. 2011, Oteng, Marra et al. 2011, Potoglou, Burge et al. 2011, Scalone, Watson et al. 2011, Sweeting, Whitty et al. 2011, Whitty, Scuffham et al. 2011) Also, during this period 3 analyses used LCM (Miguel, Ryan et al. 2005, Grindrod, Marra et al. 2010, Mentzakis, Stefanowska et al. 2011)

In Chapter 7 I applied two econometric methods (Random Effects Logit and Conditional Logit) which do not cater for preference heterogeneity before Mixed Logit and LCM results were presented. The chapter considered whether Mixed Logit or LCM provides useful information over and above models which do not cater for preference heterogeneity (Random Effects Logit and Conditional Logit).

A very large patient sample was available with enough responses completed for data analysis using all 5 econometric models (n= 863). So, I expected to be able to identify considerable preference heterogeneity. Mixed Logit found evidence of preference heterogeneity for 4 / 8 variables (wait, dep, age, and dis2), but not with respect to ‘tiss’ (closeness of tissue match). The LCM technique generated information about how coefficients vary across latent classes. We increased the number of latent classes until both the Bayesian Information Criterion and Akaike Information Criterion suggested the number of classes was optimal (i.e. when we had 4 latent classes). This indicated considerable preference heterogeneity. However, I also used Wald tests to establish whether the value of MRS for attributes varied in a statistically

significant way across the 4 classes. Results suggested MRS differed significantly for 6 / 7 attributes. However, MRS for tiss was not statistically significantly different across the 4 classes, an unexpected result. The comparison of ethnic minority vs. non-ethnic minority patient DCE preferences presented in Chapter 5 found that ethnic minority patients had statistically significantly different preferences for 'tiss', whereas non-ethnic minorities had not. So in Chapter 7 (after using Mixed Logit and LCM) I also applied Conditional Logit with dummy variables for ethnic minorities. Not unsurprisingly, this model found that preferences with respect to 'tiss; differed for ethnic minorities (MRS for 'tiss' was insignificant amongst ethnic minority patients, but highly significant for non-ethnic minority patients). This analysis demonstrates the danger of simply undertaking unthinking Mixed Logit or LCM and assuming that the information it provides on preference heterogeneity will be comprehensive. From a policy making perspective, I would argue that the most useful information about preference heterogeneity related to how preferences varied between ethnic minority and other patients. So the Conditional Logit model with interaction dummy variables for ethnic minority groups out-performed both Mixed Logit and LCM by uncovering this information about preference heterogeneity. However, this does not necessarily demonstrate that Conditional Logit with dummy variables should be used in preference to Mixed Logit or LCM. Instead, it means that researchers should probably take the analysis one step further, and use interaction dummy variables within Mixed Logit or LCM models to establish whether preferences vary for key defined sub-groups of respondents. In short, researchers still need to use their heads and develop prior hypotheses about how preferences might differ between respondent groups, because they cannot just assume that a blanket application of Mixed Logit or LCM will pinpoint key policy relevant dimensions of preference heterogeneity. This is an important message, so I am planning to submit material from this chapter for publication.

Finally, Chapter 8 of the thesis considered an issue which has not really been studied to date in DCE research (i.e. whether preferences in a DCE might be altruistically motivated and, if so, how this affects DCE results). In relation to organ donation there is a great deal of evidence that altruistic motivations are predominant (Siminoff, Mercer et al. 2007, Gill and Lowes 2008, Patel, Chadha et al. 2011). Moreover, because the issue we addressed was a priority setting one, many respondents might base their responses upon some social judgement about what might be appropriate (encompassing altruistic considerations) rather than what might be best for them. If altruistic motivations enter into respondents utility functions this is compatible with the theory of the 'caring externality' (Culyer 1976, Culyer 1980) The analysis presented in Chapter 8 utilised information obtained from a question in the patient questionnaire which asked respondents whether they had answered the questionnaire in terms of what would be best for them, what would be best for them and others, or what was best for others. Regrettably, it was not possible to obtain a sufficiently large sample of respondents who replied that they only considered what was best for them to conduct sub-group analysis for this group. So, in the end, I could only compare respondents who claimed to only consider what was best for others (pure altruists) with other patients who claimed they had preferences which were at least partly self-motivated. The renal patient questionnaire contained both VAS and Eq-5d quality of life measures to ascertain respondents' quality of life. In a series of analyses in Chapter 8, I assessed whether patients who are altruistic might register preferences which involve prioritizing people for renal transplant whose quality of life may differ considerably from their own (as measured using Eq-5d and VAS); whereas more self-interested respondents would tend to prioritize potential recipients with quality of life or length of life attributes more similar to themselves.

The analysis contained in Chapter 8 reveals a complex picture. It examines possible links between the following three aspects:

(i) Respondents' quality of life as measured using Eq-5d or VAS.
(ii) Whether respondents claimed to only consider others, or whether they admitted to being at least partly self-interested (i.e. considered themselves and others, or just considered themselves).
(iii) Respondents' preferences for the attributes relating to prioritizing recipients with diseases affecting life expectancy or diseases affecting quality of life.

Consideration of the links between i, ii and iii above for Eq-5d suggests there is some limited evidence that i above affects iii above in line with what we would expected given respondents' categories in terms of ii above. However, there is little evidence of such links between i, ii, and iii above for VAS, and findings for VAS are sometimes counter-intuitive. I am inclined to think a paper using this analysis which looks at the links between i, ii, and iii, for Eq-5d might usefully add to the literature. Moreover, as I have reflected in Chapter 8, a major limitation of the analysis was that a large enough sample of respondents who admitted to being entirely self-interested could not be identified. Therefore, for my new renal dialysis DCE I have posed a similar altruism question, and there are enough respondents in each of the three groups to enable comparison of completely self-interested patients with those claiming they are purely altruistically motivated. I am hopeful that this new DCE dataset will provide more useful data to allow consideration of how altruistic vs. self-interested motivations might affect preferences given respondents' Eq-5d and VAS status.

4. Conclusions.

This thesis has updated earlier systematic reviews relating to the use of DCEs in health. Material from Chapter 1 of the thesis could therefore underpin a future submission for

publication. Material from chapter 2 of the thesis has been published in Health Policy (Clark, Moro et al. 2009). The analysis contained in chapters 3 and 4 relating to whether respondents take differences in the monetary attribute into account, addressed an interesting question. However, to reach robust conclusions, the analysis should ideally be re-worked using a DCE with an incentive compatible monetary attribute; well informed payment vector(s); and a strong qualitative agenda using interviews and 'think aloud' exercises to establish what meaning respondents attach to the monetary attribute. The other finding of Chapter 4, that estimates of MWTP did not differ according to the monetary descriptor, is reassuring and suggests that descriptors which refer to cost might not necessarily induce 'cost-based' responses which fail to take 'consumer surplus' into account.

The analyses in chapters 7 – 8, relating to the renal transplant DCE, have produced interesting findings. Material in chapters 5, and 6, is likely to inform UK renal transplant policy when this is next re-appraised (the analysis in Chapter 5 will be published by BMC Nephrology, and much of the analysis in Chapter 6 was published in 2009 (Clark, Gumber et al. 2009). The analysis in Chapter 7 has produced an unexpected methodological result i.e. that a Conditional Logit model with dummy variables for ethnic minorities can expose differences in preferences with respect to the tissue match attribute for ethnic minority patients, when both Mixed Logit and LCM (without dummy variables) failed to highlight a difference. This highlights the need for researchers to think through how preferences might vary between respondent groups, and to consider using appropriately specified interaction dummy variables within Mixed Logit or LCM to test for hypothesized differences. This is a useful finding which I now aim to submit for publication (probably to Value in Health). Finally, Chapter 8 presents an exploration of a new issue i.e. the possible impact of altruistic preferences upon DCE responses. The analysis here, linking patient Eq-5d status to preferences under a veil of altruism, may be publishable as it demonstrates some links.

Moreover, I will continue to address issues relating to altruistic preferences, when analyzing data from my new renal dialysis DCE.

5. References.

Bech, M., T. Kjaer and J. Lauridsen (2011). "Does the Number of Choice Sets Matter? Results from a Web Survey Applying a Discrete Choice Experiment." Health Econ **20**(3): 273-286.

Bellary, S., J. P. O'Hare, N. T. Raymond, A. Gumber, S. Mughal, A. Szczepura, S. Kumar and A. H. Barnett (2008). "Enhanced diabetes care to patients of south Asian ethnic origin (the United Kingdom Asian Diabetes Study): a cluster randomised controlled trial." Lancet **371**(9626): 1769-1776.

Blaauw, D., E. Erasmus, N. Pagaiya, V. Tangcharoensathein, K. Mullei, S. Mudhune, C. Goodman, M. English and M. Lagarde (2010). "Policy interventions that attract nurses to rural areas: a multicountry discrete choice experiment." Bull World Health Organ **88**(5): 350-356.

Carlsson, F. and P. Martinsson (2001). "Do hypothetical and actual marginal willingness to pay differ in choice experiments? Application to the valuation of the environment." Journal of Environmental Economics and Management **41**(2): 179-192.

Clark, M., D. Moro and A. Szczepura (2009). "Balancing patient preferences and clinical needs: community versus hospital based care for patients with suspected DVT." Health Policy **90**(2-3): 313-319.

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

Coast, J., H. Al-Janabi, E. J. Sutton, S. A. Horrocks, A. Vosper, J., D. R. Swancutt and T. N. Flynn (2012). "Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations." Health Econ **21**(6): 730-741.

Coast, J. and S. Horrocks (2007). "Developing attributes and levels for discrete choice experiments using qualitative methods." J Health Serv Res Policy **12**(1): 25-30.

Culyer, A. J. (1976). Need and the National Health Service : economics and social choice. Totowa, N.J., Rowman and Littlefield.

Culyer, A. J. (1980). The political economy of social policy. New York, St. Martin's Press.

de Bekker-Grob, E. W., R. Hofman, B. Donkers, M. van Ballegooijen, T. J. Helmerhorst, H. Raat and I. J. Korfage (2010). "Girls' preferences for HPV vaccination: a discrete choice experiment." Vaccine **28**(41): 6692-6697.

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

Eberth, B., V. Watson, M. Ryan, J. Hughes and G. Barnett (2009). "Does one size fit all? Investigating heterogeneity in men's preferences for benign prostatic hyperplasia treatment using mixed logit analysis." Med Decis Making **29**(6): 707-715.

Gill, P. and L. Lowes (2008). "Gift exchange and organ donation: donor and recipient experiences of live related kidney transplantation." Int J Nurs Stud **45**(11): 1607-1617.

Goto, R., S. Nishimura and T. Ida (2007). "Discrete choice experiment of smoking cessation behaviour in Japan." Tobacco Control **16**(5): 336-343.

Goto, R., Y. Takahashi and T. Ida (2011). "Changes in Smokers' Attitudes Toward Intended Cessation Attempts in Japan." Value in Health **14**(5): 785-791.

Grindrod, K. A., C. A. Marra, L. Colley, R. T. Tsuyuki and L. D. Lynd (2010). "Pharmacists' preferences for providing patient-centered services: a discrete choice experiment to guide health policy." Ann Pharmacother **44**(10): 1554-1564.

Gyrd-Hansen, D. and U. S. Skjoldborg (2008). "The price proxy in discrete choice experiments: Issues of relevance for future research." Chapter 8, Ryan, M., Gerard K., and

Amaya-Amaya (eds.), Using Discrete Choice Experiments to Value Health and Health Care, 175-193.

Hall, J., D. G. Fiebig, M. T. King, I. Hossain and J. J. Louviere (2006). "What influences participation in genetic carrier testing? Results from a discrete choice experiment." J Health Econ **25**(3): 520-537.

Hauber, A. B., A. F. Mohamed, F. R. Johnson and H. Falvey (2009). "Treatment preferences and medication adherence of people with Type 2 diabetes using oral glucose-lowering agents." Diabet Med **26**(4): 416-424.

Howard, K. and G. Salkeld (2009). "Does Attribute Framing in Discrete Choice Experiments Influence Willingness to Pay? Results from a Discrete Choice Experiment in Screening for Colorectal Cancer." Value in Health **12**(2): 354-363.

Howard, K. and G. Salkeld (2009). "Does attribute framing in discrete choice experiments influence willingness to pay? Results from a discrete choice experiment in screening for colorectal cancer." Value Health **12**(2): 354-363.

Johnson, F. R. and W. H. Desvousges (1997). "Estimating stated preferences with rated-pair data: Environmental, health, and employment effects of energy programs." Journal of Environmental Economics and Management **34**(1): 79-99.

Johnson, F. R., A. F. Mohamed, S. Ozdemir, D. A. Marshall and K. A. Phillips (2011). "How Does Cost Matter in Health-Care Discrete-Choice Experiments?" Health Econ **20**(3): 323-330.

Johnson, F. R., S. Ozdemir and K. A. Phillips (2010). "Effects of simplifying choice tasks on estimates of taste heterogeneity in stated-choice surveys." Soc Sci Med **70**(2): 183-190.

Lancsar, E. and J. Louviere (2006). "Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences?" Health Econ **15**(8): 797-811.

Lancsar, E. J., J. P. Hall, M. King, P. Kenny, J. J. Louviere, D. G. Fiebig, I. Hossain, F. C. K. Thien, H. K. Reddel and C. R. Jenkins (2007). "Using discrete choice experiments to investigate subject preferences for preventive asthma medication." Respirology **12**(1): 127-136.

Mark, T. L. and J. Swait (2004). "Using stated preference and revealed preference modeling to evaluate prescribing decisions." Health Econ **13**(6): 563-573.

Mentzakis, E., P. Stefanowska and J. Hurley (2011). "A discrete choice experiment investigating preferences for funding drugs used to treat orphan diseases: an exploratory study." Health Econ Policy Law **6**(3): 405-433.

Miguel, F. S., M. Ryan and M. Amaya-Amaya (2005). "'Irrational' stated preferences: a quantitative and qualitative investigation." Health Econ **14**(3): 307-322.

Mohamed, A. F., J. D. Epstein and J. M. Li-McLeod (2011). "Patient and parent preferences for haemophilia A treatments." Haemophilia **17**(2): 209-214.

Oteng, B., F. Marra, L. D. Lynd, G. Ogilvie, D. Patrick and C. A. Marra (2011). "Evaluating societal preferences for human papillomavirus vaccine and cervical smear test screening programme." Sex Transm Infect **87**(1): 52-57.

Ozdemir, S., F. R. Johnson and A. B. Hauber (2009). "Hypothetical bias, cheap talk, and stated willingness to pay for health care." J Health Econ **28**(4): 894-901.

Patel, S. R., P. Chadha and V. Papalois (2011). "Expanding the live kidney donor pool: ethical considerations regarding altruistic donors, paired and pooled programs." Exp Clin Transplant **9**(3): 181-186.

Potoglou, D., P. Burge, T. Flynn, A. Netten, J. Malley, J. Forder and J. E. Brazier (2011). "Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data." Soc Sci Med **72**(10): 1717-1727.

Ratcliffe, J. (2000). "The use of conjoint analysis to elicit willingness-to-pay values. Proceed with caution?" Int J Technol Assess Health Care **16**(1): 270-275.

Regier, D. A., J. M. Friedman, N. Makela, M. Ryan and C. A. Marra (2009). "Valuing the benefit of diagnostic testing for genetic causes of idiopathic developmental disability: willingness to pay from families of affected children." Clin Genet **75**(6): 514-521.

Ryan, M. and K. Gerard (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., V. Watson and V. Entwistle (2009). "Rationalising the 'irrational': a think aloud study of discrete choice experiment responses." Health Econ **18**(3): 321-336.

Ryan, M. and S. Wordsworth (2000). "Sensitivity of willingness to pay estimates to the level of attributes in discrete choice experiments." Scottish Journal of Political Economy **47**(5): 504-524.

Scalone, L., V. Watson, M. Ryan, N. Kotsopoulos and R. Patel (2011). "Evaluation of patients' preferences for genital herpes treatment." Sexually Transmitted Diseases **38**(9): 802-807.

Scuffham, P. A., J. A. Whitty, M. Taylor and R. C. Saxby (2010). "Health system choice: a pilot discrete-choice experiment eliciting the preferences of British and Australian citizens." Appl Health Econ Health Policy **8**(2): 89-97.

Siminoff, L., M. B. Mercer, G. Graham and C. Burant (2007). "The reasons families donate organs for transplantation: implications for policy and practice." J Trauma **62**(4): 969-978.

Skjoldborg, U. S. and D. Gyrd-Hansen (2003). "Conjoint analysis. The cost variable: an Achilles' heel?" Health Econ **12**(6): 479-491.

Sweeting, K. R., J. A. Whitty, P. A. Scuffham and M. J. Yelland (2011). "Patient preferences for treatment of achilles tendon pain: results from a discrete-choice experiment." Patient **4**(1): 45-54.

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

van Helvoort-Postulart, D., T. van der Weijden, B. G. Dellaert, M. de Kok, M. F. von Meyenfeldt and C. D. Dirksen (2009). "Investigating the complementary value of discrete choice experiments for the evaluation of barriers and facilitators in implementation research: a questionnaire survey." Implement Sci **4**: 10.

Watson V and Ryan M (2010). Valuing Patient Experiences in the provision of Chlamydia Screening, Health Economics Research Unit Briefing Paper, March 2010. Aberdeen, Health Economics Research Unit

Whitty, J. A., P. A. Scuffham and S. R. Rundle-Thiele (2011). "Public and decision maker stated preferences for pharmaceutical subsidy decisions: a pilot study." Appl Health Econ Health Policy **9**(2): 73-79.

Wittink, M. N., M. Cary, T. Tenhave, J. Baron and J. J. Gallo (2010). "Towards Patient-Centered Care for Depression: Conjoint Methods to Tailor Treatment Based on Preferences." Patient **3**(3): 145-157.

Appendix A

Who should be prioritized for renal transplantation? Assessment of how renal patient preferences are influenced by patient characteristics.

1. Introduction

In chapter 5 I assessed the preferences of key stakeholder groups (patients, carers, live donors / relatives of deceased donors, and healthcare professionals) as well as ethnic minorities (a category including white ethnic minorities) vs. non-ethnic minority patients. Chapter 6 then addressed diversity issues within the patient group comparing preferences between non-white ethnic minority patients vs. other patients; South Asian vs. other patients; and male vs. female patients. However, in addition to the analysis in those chapters I also conducted a wealth of sub-group analysis within the patient group using Random Effects Probit with interaction dummy variables. Unlike most of the material in the main body of the PhD, this is not of particularly great methodological interest. Rather instead, it might be of interest to readers who are interested in the preferences of particular sub-groups of renal transplant patients. Therefore this material is placed in this appendix for reference purposes, rather than in the main body of the thesis.

Sub-groups relate to transplant status (whether respondents have had a successful transplant, failed transplant, or are awaiting a transplant); age; and whether respondents have dependent children or dependent adults. The issue of links between quality of life (measured both using Eq-5d, and the Visual Analogue scale [VAS]) and preferences is not explored in this appendix, because these issues are explored in detail in chapter 8 of the thesis.

2. Background.

The ‘Background’ to this research is already discussed in detail in chapter 5, section 2, so the reader is referred to that text. The other background information that readers might like to be aware of is because we had a large sample of patient responses it afforded the opportunity to conduct in-depth analysis to determine how patient preferences are affected by respondent characteristics within the patient group. We can establish how preferences vary in relation to a range of patient characteristics. This is useful because patient valuation of some attributes, may be affected by their own characteristics. The material in this appendix therefore reports upon this analysis.

3. Materials and methods.

3.1. Pilot exercise.

I refer the reader to section 3.1 of Chapter 5, which provides information about piloting of the DCE.

3.2. Selection of attributes and levels for the final DCE.

I refer the reader to section 3.2 of Chapter 5, which provides information about selection of attributes and levels for the DCE. Details of the attributes and levels used for the final analysis are presented in table 1.

Table 1: Final attributes and levels.

Attribute	Variable name	Levels	Interpretation of coefficients.
Time spent awaiting transplantion	wait	1 month, 2 years, and 10 years.	Indirect utility of each 1 year reduction in transplant recipient waiting time.
Tissue type matching	tiss	Non-favourable match: 86% average kidney survival rate post-transplant. Favourable match: 89% average kidney survival rate post-transplant. Perfect match: 90% average kidney survival rate post-transplant.	Indirect utility of prioritizing people for each 1% improvement in kidney survival.
How many child or adult dependents recipients have	dep	None, 1, or 4 dependents.	Indirect utility of each additional dependent.
Recipient age	age	20 years, 45 years, and 65 years	Indirect utility for each 1 year reduction in recipient age.
Diseases predominantly affecting life expectancy	dis1	No disease affecting life expectancy (other than Kidney disease) vs. moderate disease (uncontrolled hypertension or obesity) & Kidney disease.	Indirect utility of having no rather than moderate disease predominantly affecting life expectancy.
	dis2	Moderate disease (uncontrolled hypertension or obesity) affecting life expectancy vs. severe disease (heart attack, stroke, or diabetes with complications).	Indirect utility of having moderate disease rather than severe disease predominantly affecting life expectancy.
Diseases predominantly affecting quality of life	ill1	No disease affecting quality of life (other than Kidney disease) vs. moderate disease (mild asthma).	Indirect utility of having no disease rather than a moderate disease predominantly affecting quality of life.
	ill2	Moderate disease (mild asthma) affecting quality of life vs. severe disease (severe arthritis).	Indirect utility of having a moderate disease rather than a severe disease predominantly affecting quality of life.

3.3. Development of final DCE.

I refer the reader to section 3.3 of chapter 5, which explains how the final DCE questionnaire was developed.

3.4. Questionnaire distribution.

We included 20,000 flyers and freepost reply envelopes in Kidney Life inviting patients, carers, donors, or healthcare workers to request a questionnaire. Patient questionnaires were mainly obtained as a result of obtaining responses to that mailout. However, because we had a lack of responses from ethnic minority patients we obtained 5 additional patient responses from patients at the University Hospitals of Coventry and Warwickshire NHS Trust, and 18 via Ealing hospitals NHS Trust.

3.5. Data analysis.

We used model type A, to establish preferences for the patient group overall.

$$Y_{ij} = \beta_0 + \beta_1 wait_{ij} + \beta_2 tiss_{ij} + \beta_3 dep_{ij} + \beta_4 age_{ij} + \beta_5 dis1_{ij} + \beta_6 dis2_{ij} + \beta_7 ill1_{ij} + \beta_8 ill2_{ij} + \mu_i + \varepsilon_{ij}$$

(Model type A: model 1)

Here Y_{ij} is the binary dependent variable, from individuals i = 1…m, for observations j = 1…n_i. The number of observations n_i varies because the i individuals do not all complete every pairwise choice (a minority of respondents don't answer all choices). Whilst μ_i is the random effects error term (which allows for multiple responses from i respondents) and ε_{ij} is the

standard Probit error term for individuals i for j observations, and the other variables are as defined in table 1.

We also use model type B, to compare preferences patient preferences for subgroups of patients to other patients. Here Y_{ij} , μ_i, and ε_{ij} are as previously defined, whilst the D_s prefix indicates a dummy variable, for the sub-group in question (see table 2).

$$Y_{ij}= \beta_0+ \beta_1 wait_{ij}+\beta_2 tiss_{ij}+\beta_3 dep_{ij}+\beta_4 age_{ij}+\beta_5 dis1_{ij} +\beta_6 dis2_{ij}+\beta_7 ill1_{ij}+\beta_8 ill2_{ij}$$
$$+ \beta_9 D_s+ \beta_{10} D_{sij} wait_{ij} +\beta_{11} D_{sij} tiss_{ij}+\beta_{12} D_{sij} dep_{ij}+\beta_{13} D_{sij} age_{ij}+\beta_{14} D_{sij} dis1_{ij}$$
$$+\beta_{15} dis2_{ij}+\beta_{16} D_{sij} ill1_{ij}+\beta_{17} D_{sij} ill2_{ij}+ \mu_i +\varepsilon_{ij}$$

(Model type B: models 2, 3, 4, 6, 7, and 8)

Model type C, can be used when you want to compare the upper and lower quartiles of a continuously distributed patient attribute, such as age, with the inter-quartile range grouping. Here Y_{ij} , μ_i, and ε_{ij} are as previously defined, whilst the D_{LQ} and D_{HQ} prefixes are defined in table 2.

$$Y_{ij}=\beta_{0+}\beta_1 wait_{ij}+\beta_2 tiss_{ij}+\beta_3 dep_{ij}+\beta_4 age_{ij}+\beta_5 dis1_{ij}+\beta_6 dis2_{ij}+\beta_7 ill1_{ij}+\beta_8 ill2_{ij}$$
$$+\beta_9 D_{LQ}+\beta_{10} D_{LQij} wait_{ij}+\beta_{11} D_{LQij} tiss_{ij}+\beta_{12} D_{LQij} dep_{ij}+\beta_{13} D_{LQij} age_{ij}+\beta_{14} D_{LQij} dis1_{ij}$$
$$+\beta_{15} D_{LQij} dis2_{ij}+\beta_{16} D_{LQij} ill1_{ij} +\beta_{17} D_{LQij} ill2_{ij} +\beta_{18} D_{HQ}+ \beta_{19} D_{HQij} wait_{ij}+\beta_{20} D_{HQij} tiss_{ij}$$
$$+\beta_{21} D_{HQij} dep_{ij}+\beta_{22} D_{HQij} age_{ij}+\beta_{23} D_{HQij} dis1_{ij}+\beta_{24} D_{HQij} dis2_{ij}+\beta_{25} D_{HQij} ill1_{ij} +\beta_{26} D_{HQij} ill2_{ij}+\mu +\varepsilon$$

(Model type C: Model 5)

Table 2: Dummy variables.

Model type B	**(Models 2, 3, 4, 6, 7, & 8).**	
Respondent grouping	**Dummy variable (prefix upon variable name)**	**Coding for dummy variable**
Base group (preferences for patients not in the defined sub-group)	Not / Applicable	Not required: Preferences of those not in the sub-group correspond to the base group coefficients.
Patient sub-group preferences	D_s	D_s = 1 for the subgroup D_s = 0 otherwise
Model type C	**(Model 5).**	
Respondent grouping	**Dummy variable (prefix upon variable name)**	**Coding for dummy variable**
Base group (preferences for patients not in the defined sub-group)	Not / Applicable	Not required: Preferences of those not in the sub-groups correspond to the base group coefficients.
Those in the lowest quartile of the continuously distributed variable	D_{LQ}	D_{LQ} = 1 for respondents in the lowest quartile D_{LQ} = 0 otherwise
Those in the highest quartile of the continuously distributed variable	D_{HQ}	D_{HQ} = 1 for respondents in the highest quartile D_{HQ} = 0 otherwise

3.6. Establishing Marginal Rate of Substitution (MRS).

To express the value of changes in attributes with respect to changes in another we calculate MRS (see table 3). MRS values changes in the other variables compared with a 1 year change in waiting time. To establish whether MRS is significant we used the Delta method (Wooldridge 2002) which was executed using the 'nlcom' command in STATA v. 9.2, to derive 95% confidence intervals. We use Wald tests (executed using 'testnl in STATA) to establish whether there is a statistically significant difference in MRS between the base group and dummy variables group(s). Difference in MRS is indicated by a Wald test p-value ≤ 0.05.

Table 3: Calculating MRS.

Model type A **Variable**	**MRS all respondents**	**Model type A.** **Variable**	**MRS all respondents**
wait	N / A	**dis1**	β_5 / β_1
tiss	β_2 / β_1	**dis2**	β_6 / β_1
dep	β_3 / β_1	**ill1**	β_7 / β_1
age	β_4 / β_1	**ill2**	β_8 / β_1
Model type B **Variable**	**Base group MRS**	**Model type B** **Variable**	**Defined sub-group MRS**
wait	N / A	**wait**	N / A
tiss	β_2 / β_1	**tiss**	$(\beta_2 + \beta_{11}) / (\beta_1 + \beta_{10})$
dep	β_3 / β_1	**dep**	$(\beta_3 + \beta_{12}) / (\beta_1 + \beta_{10})$
age	β_4 / β_1	**age**	$(\beta_4 + \beta_{13}) / (\beta_1 + \beta_{10})$
dis1	β_5 / β_1	**dis1**	$(\beta_5 + \beta_{14}) / (\beta_1 + \beta_{10})$
dis2	β_6 / β_1	**dis2**	$(\beta_6 + \beta_{15}) / (\beta_1 + \beta_{10})$
ill1	β_7 / β_1	**ill1**	$(\beta_7 + \beta_{16}) / (\beta_1 + \beta_{10})$
ill2	β_8 / β_1	**ill2**	$(\beta_8 + \beta_{17}) / (\beta_1 + \beta_{10})$
Model type C **Variable**	**Base group MRS**	**Model type C** **Variable**	**Lower quartile MRS**
wait	N / A	**wait**	N / A
tiss	β_2 / β_1	**tiss**	$(\beta_2 + \beta_{11}) / (\beta_1 + \beta_{10})$
dep	β_3 / β_1	**dep**	$(\beta_3 + \beta_{12}) / (\beta_1 + \beta_{10})$
age	β_4 / β_1	**age**	$(\beta_4 + \beta_{13}) / (\beta_1 + \beta_{10})$
dis1	β_5 / β_1	**dis1**	$(\beta_5 + \beta_{14}) / (\beta_1 + \beta_{10})$
dis2	β_6 / β_1	**dis2**	$(\beta_6 + \beta_{15}) / (\beta_1 + \beta_{10})$
ill1	β_7 / β_1	**ill1**	$(\beta_7 + \beta_{16}) / (\beta_1 + \beta_{10})$
ill2	β_8 / β_1	**ill2**	$(\beta_8 + \beta_{17}) / (\beta_1 + \beta_{10})$
Model type C **Variable**	**Higher quartile MRS**		
wait			
tiss	$(\beta_2 + \beta_{20}) / (\beta_1 + \beta_{19})$		
dep	$(\beta_3 + \beta_{21}) / (\beta_1 + \beta_{19})$		
age	$(\beta_4 + \beta_{22}) / (\beta_1 + \beta_{19})$		
dis1	$(\beta_5 + \beta_{23}) / (\beta_1 + \beta_{19})$		
dis2	$(\beta_6 + \beta_{24}) / (\beta_1 + \beta_{19})$		
ill1	$(\beta_7 + \beta_{25}) / (\beta_1 + \beta_{19})$		
ill2	$(\beta_8 + \beta_{26}) / (\beta_1 + \beta_{19})$		

4. Results.

4.1. Results - Sample characteristics.

Table 4: Patient sample characteristics.

	Patients (n = 908)
AGE	
Mean age	54.88 years
GENDER	
Male	508 (55.9%)
Female	397 (43.7%)
Not indicated	3 (0.3%)
ETHNICITY	
White (British)	799 (88%)
White ethnic minorities	27 (2.9%)
Non-white ethnic minorities	69 (7.6%)
Asian groups (also included in non-Asian ethnic minority category)	50 (5.5%)
Not indicated	13 (1.4%)
DEPENDENT CHILDREN	
0	755 (83.1%)
1	72 (7.9%)
2	49 (5.4%)
3	12 (1.3%)
> 3	7 (0.8%)
Not indicated	13 (1.4%)
DEPENDENT ADULTS	
0	750 (82.6%)
1	121 (13.3%)
2	17 (1.9%)
> 2	8 (0.9%)
Not indicated	12 (1.3%)

Table 4 indicates respondent characteristics. UK Renal Registry data (Byrne, Ford et al. 2008; Byrne, Steenkamp et al. 2008) was used to assess patient sample representativeness. Of the 895/ 908 patients indicating ethnicity, 799 / 895 patients (89.3%) were white (British), and 27 / 895 (3%) were white ethnic minorities, so 92.3% are white. UK incidence data (Byrne, Ford et al. 2008) suggested 79.7% of renal patients are white, so whites are over-

represented in our survey. Overall, 69 / 895 (7.7%) patients indicating ethnicity were non-white, compared with a 20.3% incidence rate (Byrne, Ford et al. 2008), 50 / 69 non-white patients were South Asians (5.6% of those indicating ethnicity) compared to a 10.5% incidence (Byrne, Ford et al. 2008). 508 /908 patients (55.9%) were male, 397 / 908 (43.7%) were female, 3 / 908 (0.3%) did not say. Graphically presented Renal Registry data (Byrne, Ford et al. 2008) reassuringly indicated slightly higher proportions of male than female patients across age groups. Average sample patient age was 54.88 years (median 57 years), and Renal Registry data median age (57.3 years) was virtually identical (Byrne, Steenkamp et al. 2008).

The patient sample comprised: 468 / 908 (51.5%) with successful transplants; 118 / 908 (13%) whose transplant failed; 279 / 908 (30.7%) awaiting transplants (average wait 22.6 months). Some patients whose transplant failed also appeared as awaiting transplantation; 237 / 908 (26.3%) had dialysis without transplantation; and 57 / 908 (6.3%) had kidney disease, not requiring dialysis. Renal Registry prevalence data (Byrne, Steenkamp et al. 2008) suggests 46.9% of patients have successful transplants (close to our figure). There is no incidence / prevalence data for other patient categories. Amongst non-whites (including Asians) figures are 18 / 69 patients (26%) with successful transplants; 10 / 69 (14.5%) whose transplant failed; 35 / 69 patients (50.7%) awaiting a transplant on dialysis (average wait: 21.45 months); and 3 / 69 (4.3%) with kidney disease, not requiring dialysis. Unfortunately renal registry data (Byrne, Steenkamp et al. 2008) does not indicate ethnicity. However, lower percentage figures for successes, and higher transplant failures figures are expected (ethnic minorities are less likely to be closely matched).

4. Results – Econometric analysis.

4.2. Econometric analysis of patient preferences overall.

Table 5 - Model 1: Patient preferences.

Attribute	Coefficient	Implied MRS – relative to waiting time	Attribute	Coefficient	Implied MRS – relative to waiting time
wait	0.444**	1	**dis1**	-.0005	-0.01 (-1.03 / 1.01)
tiss	.0626**	1.41** (1.08 / 1.74)	**dis2**	.6812**	15.33** (13.46 / 17.20)
dep	.0636**	1.43** (1.17 / 1.69)	**ill1**	-.1213**	-2.73** (-1.45 / -4.00)
age	.0069**	0.16** (0.12 / 0.19)	**ill2**	.1857**	4.18** (3.12 / 5.24)
Constant	.1225**		**Log-likelihood:**	-5609.41	
Number of respondents:	908	**% of actual values predicted:**	62.82%	**Mc Faddens R^2:**	0.110

*: Denotes significant at 5% level; **: Denotes significant at 1% level

Results for patients overall (model 1 – table 5) are reassuring. The model accurately predicts 62.82% of responses and McFaddens R^2 = 0.110. Overall 7 / 8 variables are significant at the 1% level, but one (dis1) is insignificant.

The variable tiss has MRS = 1.41 [CI: 1.08 / 1.74]. Implying that the MRS of transplanting to someone with a favourable rather than non-favourable match is equivalent to 4.23 because there is a 3% difference in graft survival. Therefore if you had 2 patients competing for transplantation of one kidney (one with a non-favourable match, and the other a favourable match), then all other things being equal, you would expect the patient with the less favourable match to have waited 4.23 years longer than the other patient to be of equal priority, and more than 4.23 years to be a greater priority. If the comparison is between 2 patients but one had a non-favourable match, and the other a perfect match, it would require

the non-favourable match to have waited 5.84 years to be of equal priority (assuming other characteristics are identical), because there is a 4% difference in the likelihood of kidney survival. Likewise, if the difference between them was a favourable vs. a perfect match, it would require the non-favourable match to wait 1.41 years longer, to be of equal priority (because there is a 1% difference in transplant survival).

The variable dep has MRS = 1.43 [CI: 1.17 / 1.69]. Someone with no dependents (all other things being equal) would be expected to wait an extra 1.43 years for a transplant, to be considered an equal priority to someone with 1 dependent.

MRS for age, indicates patient valuation of prioritising younger rather than older patients (for each year younger). At 0.16 [0.12 / 0.19], it suggests that if you had two patients in competition for one organ, and the older one had waited 1 year longer than the younger one, all other things being equal the younger one would have to have be 6.25 years younger (i.e. 1 / 0.16 = 6.25) to be of equal priority, and wait more than 6.25 years to be a greater priority.

Patients don't prioritize people without a disease affecting life expectancy more than those with a moderate disease affecting (dis1). They would prioritize those with a moderate disease affecting life expectancy, rather than a severe disease, and MRS = 15.33 [CI: 13.46 / 17.20]. This implies all other things being equal, a person with a severe disease affecting life expectancy not a moderate one, would have to wait 15.33 years longer to be of equal priority to a person with a moderate disease affecting life expectancy.

The coefficient on ill1 is negative, suggesting patients would prioritize someone with a moderate disease affecting quality of life, to a person with no disease (other than kidney disease), and MRS is -2.73 [-1.45 / -4.00]. Thus all other things being equal, a person without

a disease affecting quality of life would be expected to wait 2.73 years longer, to be of equal priority. This result, is probably because many patient respondents have moderate diseases affecting quality of life.

The coefficient on ill2 is positive, people would prioritize a person with a moderate disease affecting quality of life more than someone with severe disease. With MRS = 4.18 [3.12 / 5.24], it implies that if there are 2 patients competing for transplantation of one kidney (differing only in terms of how long they had waited, and diseases affecting quality of life) then a person with a severe disease affecting quality of life, would have to have waited 4.18 years longer to be of equal priority, and more than 4.18 years to be a greater priority, than someone with moderate disease.

4.3. Econometric analysis - Patient sample, with interaction dummy variables for those who are successfully transplanted.

Table 6 - Model 2: Patients values with interaction dummy variables for those who are successfully transplanted.

Attribute	Coefficient for patients in general	Implied MRS for patients in general	Coefficient for dummy variable for successfully transplated patients.	Implied MRS for the successfully transplanted group	Wald test p-values
wait	.0400**	1	.0092*	1	
tiss	.0574**	1.43** (0.91 / 1.96)	.0089	1.35** (0.92 / 1.77)	p=0.0310
dep	.0640**	1.60** (1.18 / 2.02)	.0002	1.31** (0.98 / 1.63)	p=0.0004
age	.0061**	0.15** (0.10 / 0.21)	.0011	0.15** (0.10 / 0.19)	p=0.0583
dis1	.0110	0.28 (-1.34 / 1.89)	-.0314	-0.41 (-1.73 / 0.90)	p=0.5664
dis2	.5921**	14.80** (11.86 / 17.73)	.1767**	15.62** (13.21 / 18.03)	p=0.0000
ill1	-.0673	-1.68 (-3.73 / 0.37)	-.1149	-3.70** (-2.07 / -5.33)	p=0.6032
ill2	.1134**	2.83** (1.14 / 4.53)	.1466**	5.28** (3.92 / 6.64)	p=0.6373
Intercepts	.0905**		.0649*		
% of actual values predicted:	63.12%	**Sample:**	896 patients (468 had been successfully transplanted).	**McFaddens R^2:**	0.113
LR test (λ):	21.64	**Jointly significant?:**	Yes: CV for 9 df = 16.92	**Log-likelihood:**	-4919.38

Model 2 (table 6) establishes whether preferences differ amongst patients who report they have been successfully transplanted (compared to patients overall). The model accurately predicts 63.12% of values and McFaddens $R^2 = 0.113$. The likelihood ratio (LR) test suggests that the dummy variables are jointly significant ($\lambda = 21.64$, and the critical value (CV) for 9 degrees of freedom [df] is 16.92). There is evidence of a statistically significant difference in 'impacts' with respect to 3 variables (wait, dis2, and ill2). Clearly because wait is used to derive MRS, there is not a figure for MRS for wait, however the dummy variable is positive

suggesting that those who are successfully transplanted prioritize those who have been waiting by more than those patients who have not.

For dis2 there is a statistically significant difference in MRS, according to the Wald test (p=0.0000). MRS for dis2 (MRS = 15.62 [CI: 13.21 / 18.03] vs. 14.80 [CI: 11.86 / 17.73]), suggesting those who have been successfully transplanted place higher upon transplanting to those with moderate rather than severe disease affecting life expectancy. This makes sense, those who have been successfully transplanted are unlikely to have severe diseases affecting life expectancy (as if they had, they are not likely to have been offered a transplant). The dummy variable dis2 is significant and positive, suggesting that the attribute impact is higher for those who are successfully transplanted. However, surprisingly the Wald test does not suggest MRS differs between the groups.

Wald tests also suggest that MRS varies in relation to tiss (p= 0.0310) and dep (p=0.0004). For tiss (MRS = 1.35 [CI: 0.92 / 1.77] vs. 1.43 [CI: 0.91 / 1.96]), MRS point estimates suggest that those who have been successfully transplanted place a marginally lower valuation upon having a good HLA tissue match than the sample overall. For dep (MRS = 1.31 [CI: 0.98 / 1.63] vs. 1.60 [CI: 1.18 / 2.02]), MRS suggests those who are successfully transplanted, do not value prioritising those with child or adult dependents quite as much as those who are not.

4.4. Econometric analysis - Patient sample, with interaction dummy variables for those who had failed transplants.

Table 7 - Model 3: Patients values with interaction dummy variables for those who had a failed transplant.

Attribute	Coefficient for patients in general	Implied MRS for patients in general	Coefficient for dummy variable for the patients who are currently waiting for a transplant	Implied MRS for the waiting for a transplant group	Wald test p-values
wait	.0440**	1	.0054	1	
tiss	.0599**	1.36** (1.00 / 1.72)	.0177	1.57** (0.72 / 2.43)	p = 0.0920
dep	.0651**	1.48** (1.19 / 1.76)	-.0089	1.14** (0.50 / 1.78)	p = 0.0003
age	.0066**	0.15** (0.11 / 0.19)	.0012	0.16** (0.06 / 0.25)	p = 0.0627
dis1	-.0137	-0.31 (-1.43 / 0.80)	.0788	1.32 (-1.24 / 3.88)	p = 0.2596
dis2	.6781**	15.41** (13.36 / 17.45)	.0572	14.88** (10.26 / 19.51)	p = 0.0000
ill1	-.1239**	-2.81** (-1.42 / -4.20)	-.0010	-2.53 (-5.75 / 0.69)	p= 0.2622
ill2	.1970**	4.48** (3.32 / 5.63)	-.0648	2.68* (0.01 / 5.35)	p = 0.0036
Intercepts	.1244		-.0027		
% of actual values predicted:	63.66%	**Sample:**	897 (118 transplants failed	**McFaddens R^2:**	0.107
LR test (λ):	7.21	**Jointly significant?:**	No: CV for 9 df = 16.92	**Log-likelihood:**	-4930.97

Model 3 (table 7), considers whether those who have had a failed transplant have different preferences to other patients. The model accurately predicts 63.66% of values and McFaddens R^2 = 0.107. The LR test suggests that the dummy variables are not jointly significant (λ = 7.21, and the CV for 9 df is 16.92). Moreover, there is no evidence that any of the interaction dummy variables are significant at the 5% level, so variable 'impacts' do not differ significantly between the groups. However, 3 / 7 of the Wald tests do suggest that MRS

might differ. These differences apply in relation to dep (p=0.0003); dis2 (p=0.0000); and ill2 (p=0.0036).

For dep (MRS = 1.14 [CI: 0.50 / 1.78] vs. 1.48 [CI: 1.19 / 1.76]), and there a difference in MRS according to the Wald test (p=0.0003) suggesting those with a failed transplant patient value prioritizing those with child or adult dependents less. In relation to dis2 (MRS = 14.88 [CI: 10.26 / 19.51] vs. 15.41 [CI: 13.36 / 17.45]). Although the Wald test (p=0.0000) suggests a difference, the difference is very small. With respect to ill2 (MRS = 2.68 [CI: 0.01 / 5.35] vs. 4.48 [CI: 3.32 / 5.63]), suggesting those with failed transplants would not prioritize those with moderate rather than severe diseases affecting quality of life as much, which may be due to self-interest as those with failed transplants may be more susceptible to severe diseases affecting quality of life. MRS does not vary significantly for other variables.

4.5. Econometric analysis - Patient sample, with interaction dummy variables for those who are waiting for a transplant.

Table 8 - Model 4: Patients values with interaction dummy variables for those who are waiting for a transplant.

Attribute	Coefficient for patients in general	Implied MRS for patients in general	Coefficient for dummy variable for the patients who are currently waiting for a transplant	Implied MRS for the waiting for a transplant group	Wald test p-values
wait	.0451**	1	-.0003	1	
tiss	.0610**	1.35** (0.89 / 1.81)	.0007	1.38** (0.78 / 1.97)	p=0.0100
dep	.0570**	1.26** (0.91 / 1.61)	.0035	1.35** (0.89 / 1.81)	p=0.0048
age	.0073**	0.16** (0.11 / 0.21)	-.0001	0.16** (0.09 / 0.23)	p=0.0046
dis1	-.0265	-0.59 (-2.01 / 0.83)	.0466	0.45 (-1.37 / 2.27)	p=0.3481
dis2	.7081**	15.68** (13.06 / 18.31)	-.0535	14.59** (11.33 / 17.85)	p=0.0000
ill1	-.1074*	-2.38** (-0.59 / -4.17)	.0147	-2.07 (-4.36 / 0.23)	p=0.2315
ill2	.1717**	3.80** (2.33 / 5.28)	-.0115	3.57** (1.69 / 5.46)	p=0.0308
Intercepts	.1206**		-.0094		
% of actual values predicted:	62.93%	**Sample:**	732 (279 currently waiting for a transplant)	**McFaddens R^2:**	0.111
LR test (λ):	1.84	**Jointly significant?:**	No: CV for 9 df = 16.92	**Log-likelihood:**	-4029.28

Model 4 (table 8), considers whether those currently waiting for a transplant have different preferences to other patients who addressed the question about whether they are currently waiting for a transplant (n=732). The model accurately predicts 62.93% of values, and McFaddens $R^2 = 0.111$. The LR test suggests that the dummy variables for those currently waiting for a transplant are not jointly significant ($\lambda = 1.84$, and the CV for 9 df is 16.92). Moreover, there is no evidence any of the interaction dummy variables are significant (5%

level), so variable ‘impacts’ do not differ significantly between groups. However, surprisingly 5 / 7 of the Wald tests do suggest that MRS might differ. These differences apply in relation to tiss (p=0.0100); dep (p=0.0048); age (p=0.0046); dis2 (p=0.0000); and ill2 (p=0.0308).

With respect to tiss (MRS = 1.38 [CI: 0.78 / 1.97] vs. 1.35 [CI: 0.89 / 1.81]), so although Wald tests suggest MRS differs (p=0.0100), there is little difference in MRS between the groups for this variable. For dep (MRS = 1.35 [CI: 0.89 / 1.81] vs. 1.26 [CI: 0.91 / 1.61]), there is again a difference in MRS according to the Wald test (p=0.0048) but it is very marginal. In relation to age (MRS = 0.16 [CI: 0.09 / 0.23] vs. 0.16 [CI: 0.11 / 0.21]), despite the fact point estimates are equal, the Wald test suggests MRS may vary between the groups (p=0.0046), this result may have arisen because 95% confidence intervals around MRS are wider amongst those waiting for a transplant.

With respect to dis2 (MRS = 14.59 [CI: 11.33 / 17.85] vs. 15.68 [CI: 13.06 / 18.31]), so those currently awaiting a transplant may value prioritizing those with moderate rather than severe diseases affecting life expectancy very marginally less than those who are not currently waiting according to the Wald test (p=0.0000). Finally, the for the variable ill2 (MRS = 3.57 [CI: 1.69 / 5.46] vs. 3.80 [CI: 2.33 / 5.28]), the Wald test again suggests that MRS may differ between the groups (p=0.0308). However, the point estimates for MRS do not differ much. MRS for other variables does not vary significantly.

4.6. Econometric analysis - Age and patient response

Table 9 – model 5: Patients with dummy variables for those in 1st & 4th age quartiles.

Attribute	Coefficient for inter-quartile range	Implied MRS for inter-quartile range	Coefficient on dummy for patients in the lowest quartile	Implied MRS for lowest quartile	Wald test p-values
wait	.0477**	1	.0038	1	
tiss	.0608**	1.27** (0.84 / 1.71)	.0098	1.37** (0.78 / 1.96)	p=0.0392
dep	.0692**	1.45** (1.11 / 1.79)	-.0061	1.22** (0.77 / 1.67)	p=0.0002
age	.0071**	0.15** (0.10 / 0.20)	.0003	0.14** (0.08 / 0.21)	p=0.0160
dis1	-.0100	-0.21 (-1.57 / 1.15)	.1334*	2.39** (0.56 / 4.22)	p=0.0800
dis2	.6989**	14.64** (12.22 / 17.06)	-.0047	13.46** (10.37 / 16.55)	p=0.0000
ill1	-.1675**	-3.51** (-1.82 / -5.20)	-.0076	-3.40** (-1.15 / -5.64)	p=0.1393
ill2	.2405**	5.04** (3.62 / 6.46)	.0258	5.16** (3.28 / 7.04)	p=0.0174
Intercepts	.1488**		.0007		
Attribute	**Coefficient for inter-quartile range**	**Implied MRS for inter-quartile range**	**Coefficient on dummy for patients in the highest quartile**	**Implied MRS for highest quartile**	**Wald test p-values**
wait	.0477**	1	-.0154**	1	
tiss	.0608**	1.27** (0.84 / 1.71)	-.0014	1.84** (0.86 / 2.81)	p=0.0095
dep	.0692**	1.45** (1.11 / 1.79)	-.0120	1.77** (1.02 / 2.52)	p=0.0000
age	.0071**	0.15** (0.10 / 0.20)	-.0011	0.19** (0.08 / 0.29)	p=0.0008
dis1	-.0100	-0.21 (-1.57 / 1.15)	-.0890	-3.06* (-0.08 /- 6.05)	p=0.3100
dis2	.6989**	14.64** (12.22 / 17.06)	-.0594	19.80** (13.69 / 25.90)	p=0.0000
ill1	-.1675**	-3.51** (-1.82 / -5.20)	.1694*	0.06 (-3.63 / 3.75)	p=0.0011
ill2	.2405**	5.04** (3.62 / 6.46)	-.2220**	0.57 (-2.46 / 3.60)	p=0.0000
Intercepts	.1488**		-.107**		
% of actual values predicted:	62.61%	**Sample:**	889 patients: 222 in first quartile and 222 in top quartile.	**McFaddens R^2:**	0.114
LR test (λ):	36.56	**Jointly significant?:**	Yes : CV for 18 df = 28.87	**Log-likelihood:**	-4877.33

Model 5 (table 9), compares preferences for those in the lower and upper quartiles of the age distribution with those in the inter-quartile range. The model accurately predicts 62.61% of values, and McFaddens $R^2 = 0.114$. The LR test ($\lambda = 36.56$, compares with the CV of 18 df of 28.87), suggesting that the dummy variables are jointly significant. There is evidence that only dis1 is statistically significantly different at the 5% level for those in the lowest quartile of the age distribution compared with the inter-quartile range. However the Wald test (p=0.0800) suggests that MRS does not differ with respect to this variable at the 5% level. The attribute impact for the lowest quartile age grouping is positive and significant however, which contrasts with a negative and insignificant coefficient for the inter-quartile range.

The Wald tests do suggest however, that MRS for the lowest quartile group differs with respect to tiss (p=0.0392); dep (p=0.0002); age (p=0.0160) dis2 (p=0.0000), and ill2 (p=0.0174).

For tiss MRS (MRS = 1.37 [CI: 0.78 / 1.96] vs. 1.27 [CI: 0.84 / 1.71]) is significantly different between the groups (p=0.0392), and higher for those in the lowest quartile of the age distribution, implying they would prioritize those with a good tissue match more. With respect to dep, the Wald test (p=0.0002) suggests that MRS differs, and MRS (MRS = 1.22 [CI: 0.77 / 1.67] vs. 1.45 [CI: 1.11 / 1.79]) for those in the lowest quartile of the age distribution is lower, implying this group does not value prioritizing those with dependents as much as those in the inter-quartile range of the age distribution. For the attribute age the Wald test (p=0.0160) again suggests that MRS (MRS = 0.14 [CI: 0.08 / 0.21] vs. 0.15 [CI: 0.10 / 0.20]) differs. This suggests those in the lowest quartile of the age distribution might value prioritizing younger potential recipients marginally less than those in the inter-quartile range.
The Wald test on dis2 (p=0.0000) suggests that MRS (MRS = 13.46 [CI: 10.37 / 16.55] vs. 14.64 [CI: 12.22 / 17.06]) does differ, those in the lowest quartile of the age distribution

prioritize those with moderate rather than severe diseases affecting life expectancy by less than those in the inter-quartile range. Also the Wald test (p=0.0174) suggests that MRS may differ with respect to ill2 (MRS = 5.16 [CI: 3.28 / 7.04] vs. 5.04 [CI: 3.62 / 6.46]), with those in the lowest quartile of the age distribution valuing prioritizing those with moderate rather than severe diseases affecting life expectancy slightly more than those in the inter-quartile range.

Overall, 3 / 8 of the interaction dummy variables are significant at the 5% significance level for dummy variables for those in the highest quartile of the age distribution including wait, ill1 and ill2. Of course wait is the variable used to calculate MRS, so tests for differences in MRS cannot apply to it. However, it would appear that the significant and negative dummy variable on wait suggests that those in the highest quartile of the age distribution do not value prioritizing people based upon time spent on the transplant list as much.

Wald tests also suggest that ill1 (p=0.0011) and ill2 (p=0.0000) do differ between the groups for those in the highest quartile of the age distribution. For ill1 MRS (MRS = 0.006 [CI: -3.63 / 3.75] vs. -3.51 [CI: -1.82 / -5.20]) may differ, it is actually significant and negative for patients in the inter-quartile range of the age distribution, but could be positive for those in the highest quartile of the age distribution (although this is unclear because the positive point estimate is not significantly different from zero). The negative point estimate for those within the inter-quartile range of the age distribution implies respondents would prioritize those with moderate as opposed to no diseases affecting quality of life, perhaps because many respondents themselves have such diseases.

With respect to ill2, the Wald test (p=0.0000) suggests that MRS (MRS = 0.57 [CI: -2.46 / 3.60] vs. 5.04 [CI: 3.62 / 6.46]) differs between those in the upper quartile of the age distribution, and the inter-quartile range. Those in the inter-quartile range of the age

distribution would prioritize those with moderate rather than severe diseases affecting quality of life, whereas for those in the highest quartile of the age distribution, it is unclear whether they would or not.

The Wald tests also suggest that MRS varies with respect to tiss (p=0.0095); dep (p=0.0000); age (p=0.0008); and dis2 (p=0.0000), ill1 (p=0.0011) and ill2 (p=0.0000) when comparing the highest quartile of the age distribution with the inter-quartile range. For tiss, the Wald test (p=0.0095) suggests there is a difference in MRS (MRS = 1.84 [CI: 0.86 / 2.81] vs. 1.27 [CI: 0.84 / 1.71. This suggests that those in the upper quartile of the age distribution value allocating kidneys to people with a better tissue match more highly that those in the inter-quartile range. With respect to dep, the Wald test (p=0.0000) suggests that MRS (MRS = 1.77 [CI: 1.02 / 2.52] vs. 1.45 [CI: 1.11 / 1.79]) differs between the groups, with those in the highest quartile of the age distribution, placing more priority upon those with child or adult dependents. Also MRS for age (MRS = 0.19 [CI: 0.08 / 0.29] vs. 0.15 [CI: 0.10 / 0.20]) may differ between the groups (p=0.0008), with older respondents in the highest quartile of the age distribution paradoxically valuing prioritizing the young, more than the younger respondents within the inter-quartile range (so perhaps some altruism is underpinning these responses). The Wald test for dis2 (p=0.0000), suggests that MRS differs (MRS = 19.80 [CI: 13.69 / 25.90] vs. 14.64 [CI: 12.22 / 17.06]). Note Wald tests do not suggest MRS is significantly different for dis1 (p=0.3100). However, Wald tests do suggest differences for ill1 (MRS = 0.06 [CI:-3.63 / 3.75] vs. -3.51 [CI: -1.82/-5.20]) suggesting the old may value prioritizing those with no rather than moderate diseases affecting life expectancy, when the inter-quartile range does not. Also Wald tests suggest a difference in ill2 (MRS = 0.57 [CI: -2.46 / 3.60] vs. 5.04 [3.62 / 6.46]) suggesting the old would not prioritize those with moderate rather than severe diseases affecting life expectancy as much.

4.7. Econometric analysis - Comparing those with above and below average age.

Table 10 – Model 6: All patients providing details of their age, interaction dummies for above average age.

Attribute	Coefficient for patients in general	Implied MRS for patients in general	Coefficient for dummy variable for the patients with above average age	Implied MRS for those with above average age	Wald test p-values
wait	.0487**		-.0070		
tiss	.0657**	1.35** (0.89 / 1.81)	-.0055	1.45** (0.97 / 1.92)	p=0.0024
dep	.0637**	1.31** (0.95 / 1.67)	.0020	1.58** (1.20 / 1.96)	p=0.0011
age	.0074**	0.15** (0.10 / 0.20)	-.0009	0.16** (0.10 / 0.21)	p=0.0010
dis1	.0740*	1.52* (0.08 / 2.96)	-.1311**	-1.37 (-2.86 / 0.11)	p=0.0082
dis2	.7216**	14.83** (12.25 / 17.40)	-.0730	15.58** (12.87 / 18.30)	p=0.0000
ill1	-.1636**	-3.36** (-1.58 / -5.14)	.0668	-2.33* (-0.48 / -4.17)	p=0.0209
ill2	.2595**	5.33** (3.84 / 6.82)	-.1245*	3.24** (1.72 / 4.77)	p=0.0000
Intercepts	.1534**		-.0573		
% of actual values predicted:	62.49%	**Sample:**	889 patients of which 492 form the above average age group	**McFaddens R^2:**	0.113
LR test (λ):	22.47	**Jointly significant?:**	Yes: CV for 9 df = 16.92	**Log-likelihood:**	-4884.38

Model 6 (table 10) looks at whether being above or below average age affects your preferences for renal transplantation. The model accurately predicts 62.49% of responses. McFaddens $R^2 = 0.113$. The LR tests suggest that the dummy variables for above average age are jointly significant ($\lambda = 22.47$, when the CV for 9 df is 16.92). Overall 2 / 8 of the interaction dummy variables (for dis1; and ill2) are significantly different from zero at the 5% level, for the above average age group. The Wald test (p=0.0082) also suggests that MRS (MRS = -1.37 [-2.86 / 0.11] vs. 1.52 [0.08 / 2.96]) differs between the 2 groups for dis1. MRS is positive and clearly significant for the below average age base group, who would prioritize

those with no rather than moderate diseases affecting life expectancy. However, it is insignificant for those with an above average age. For ill2, the Wald test (p=0.0000) suggests that MRS differs. The MRS (MRS = 3.24 [1.72 / 4.77] vs. 5.33 [3.84 / 6.82]) is lower for those with above average age, implying a lower priority for prioritizing those with moderate rather than severe diseases affecting quality of life.

For this analysis all the Wald statistics for every variable have a p-value < 0.05 suggesting that MRS differs at the 5% level. With respect to tiss, the Wald test (p=0.0024) suggests that MRS (MRS = 1.45 [CI: 0.97 / 1.92] vs. 1.35 [CI: 0.89 / 1.81]) suggesting that those with above average age afford a marginally greater priority to prioritizing those with better tissue matches. For dep, the Wald test (p=0.0011) suggests that MRS varies between the 2 groups (MRS = 1.58 [CI: 1.20 / 1.96] vs. 1.31 [CI: 0.95 / 1.67]), with those with above average age prioritizing those with dependents more than those who have a below average age. In relation to age, the Wald test (p=0.0010) suggests that MRS may vary. However the point estimate for MRS is only slightly different (MRS = 0.16 [CI: 0.10 / 0.21] vs. 0.15 [CI: 0.10 / 0.20]), suggesting a slightly increased preference for prioritizing younger recipients, amongst those with an above average age.

With respect to dis1, the Wald test (p=0.0082) again suggests that MRS differs (MRS = -1.37 [CI: -2.86 / 0.11] vs. 1.52 [CI: 0.08 / 2.96]), with those with above average age perhaps prioritizing those with no rather than moderate diseases affecting life expectancy whilst those with below average age may have a preference in the other direction.

With respect to dis2, the Wald test (p=0.0000) again suggests that MRS differs (MRS = 15.58 [CI: 12.87 / 18.30] vs. 14.83 [CI: 12.25 / 17.40]), with those with above average age prioritizing those with moderate rather than severe diseases affecting life expectancy by

slightly more than below average age patients. There is also evidence from the Wald test (p=0.0209) that ill1 varies, and the MRS (MRS = -2.33 [CI: -0.48 / -4.17] vs. -3.36 [CI: -1.58 / -5.14]) implies that those with above average age, prioritize those with moderate rather than no diseases affecting quality of life by less than those with below average age. Also ill2 may vary and MRS (MRS = 3.24 [CI: 1.72 / 4.77] vs. 5.33 [CI: 3.84 / 6.82]) suggesting those with moderate rather than no diseases affecting quality of life are prioritized less by the old.

4.8. Econometric analysis – Comparing those with and without dependent children.

Table 11 – Model 7: Patients with interaction dummy variables for sub-group with dependent children.

Attribute	**Coefficient for patients in general**	**Implied MRS for patients in general**	**Coefficient for dummy variable for the patients who have dependent children**	**Implied MRS for the patients who have dependent children**	**Wald test** **p-values**
wait	.0430**	1	.0117*	1	
tiss	.0655**	1.52** (1.14 / 1.90)	-.0158	0.91** (0.25 / 1.57)	p=0.0005
dep	.0666**	1.55** (1.25 / 1.85)	-.0196	0.86** (0.35 / 1.37)	p=0.0000
age	.0073**	0.17** (0.13 / 0.21)	-.0019	0.10** (0.02 / 0.17)	p=0.0009
dis1	-.0066	-0.15 (-1.30 / 0.99)	.0289	0.41 (-1.75 / 2.57)	p=0.6477
dis2	.6767**	15.74** (13.60 / 17.88)	.0440	13.18** (9.56 / 16.81)	p=0.0000
ill1	-.1198**	-2.79** (-1.35 / -4.22)	-.0117	-2.40 (-5.10 / 0.29)	p=0.3035
ill2	.1838**	4.28** (3.08 / 5.47)	.0105	3.55* (1.36 / 5.75)	p=0.0461
Intercepts	.1248**		-.0159		
% of actual values predicted:	61.28%	**Sample:**	901 patients - 142 of which have dependent children	**Mc Faddens R^2:**	0.112
LR test (λ):	13.35	**Jointly significant?:**	Yes: CV for 9 df = 16.92	**Log-likelihood:**	-4956.20

Model 7 (table 11) looks at whether those with dependent children differ in their preferences. The model accurately predicts 61.28% of responses, and McFaddens R^2 = 0.112. According to the LR test the dummy variables are not jointly significant (λ = 13.35, the CV for 9 df is 16.92). We thought it was possible that those with dependents (either children or adults) might value attributes differently compared to those without dependents. The only interaction dummy variable which is significant is wait (for which MRS is not derived). The dummy variable is positive, implying that those with dependent children prioritize those who have waited longer by more than patient respondents without dependent children.

However Wald tests suggest that MRS differs with respect to tiss (p=0.0005); dep (p=0.0000); age (p=0.0009); dis2 (p=0.0000); and ill2 (p=0.0461).For tiss the Wald test (p=0.0005) suggests that MRS differs (MRS = 0.91 [CI: 0.25 / 1.57] vs. 1.52 [CI: 1.14 / 1.90]), and that those with dependent children would not prioritize those with better tissue matches as much as other patients.The Wald test (p=0.0009) suggests that those with dependent children have a different MRS for age (MRS = 0.10 [CI: 0.02 / 0.17] vs. 0.17 [CI: 0.13 / 0.21]) and surprisingly those with dependent children would prioritize transplants for younger recipients less than those without.. The Wald test for dis2 (p=0.0000) suggests MRS (MRS = 13.18 [CI: 9.56 / 16.81] vs. 15.74 [CI: 13.60 / 17.88]) differs and those with dependent children would not prioritize those with moderate rather than severe diseases affecting life expectancy as much as other patients. A Wald test (p=0.0461) also indicates that MRS may differ for ill2 (MRS = 3.55 [CI: 1.36 / 5.75] vs. 4.28 [CI: 3.08 / 5.47]), and respondents with dependent children would not prioritize those with moderate rather than severe diseases affecting quality of life as much as other patients. Wald tests for other variables are insignificant.

4.9. Econometric analysis – Comparing those with and without dependent adults.

Table 12 – model 8: Patients with interaction dummy variables for sub-group with dependent adults.

Attribute	Coefficient for patients in general	Implied MRS for patients in general	Coefficient for dummy variable for the patients who have dependent adults	Implied MRS for the patients who have dependent adults	Wald test p-values
wait	.0459**	1	-.0052	1	
tiss	.0624**	1.36** (1.01 / 1.71)	-.0053	1.40** (0.50 / 2.30)	p=0.0031
dep	.0642**	1.40** (1.12 / 1.67)	.0002	1.58** (0.86 / 2.30)	p=0.0007
age	.0067**	0.15** (0.11 / 0.18)	.0012	0.19** (0.09 / 0.30)	p=0.0277
dis1	.0037	0.08 (-1.00 / 1.17)	-.0393	-0.87 (-3.68 / 1.93)	p=05706
dis2	.6777**	14.77** (12.82 / 16.71)	.0287	17.34** (11.86 / 22.8)	p=0.0000
ill1	-.1300**	-2.83** (-1.48 / -4.19)	.0252	-2.57 (-6.05 / 0.90)	p=0.1289
ill2	.1860**	4.05** (2.93 / 5.18)	.0255	5.19** (2.30 / 8.09)	p=0.0636
Intercepts	.1306**		-.0437		
% of actual values predicted:	63.15%	**Sample:**	898 patients - 146 of which have dependent adults	**Mc Faddens R^2:**	0.112
LR test (λ):	3.97	**Jointly significant?:**	No : CV for 9 df = 16.92	**Log-likelihood:**	-4940.23

Model 8 (table 12) looks at whether those with dependent adults exhibit differences in preferences to those who say they do not. The model accurately predicts 63.15% of responses, and McFaddens R^2 = 0.112. The LR test (λ=3.97, compares with a CV for 9 df of 16.92) does not suggest that the dummy variables are jointly significant. None of the interaction dummy variables for the model which compares those with dependent adults to other patients, are statistically significantly different from zero at the 5% level. However, the Wald tests suggest that MRS may vary in relation to tiss (p=0.0031); dep (p=0.0007); age (p=0.0277); and dis2 (p=0.0000).

The Wald test for tiss (p=0.0031) suggests MRS may differ (MRS = 1.40 [CI: 0.50 / 2.30] vs. 1.36 [CI: 1.01 / 1.71]) with those with adult dependents putting marginally more emphasis upon prioritizing those with better tissue matches. For dep the Wald test (p=0.0007) suggests that MRS (MRS = 1.58 [CI: 0.86 / 2.30] vs. 1.40 [CI: 1.12 / 1.67]) may differ, with those with adult dependents wanting to prioritize those with dependents marginally more (which makes intuitive sense). The Wald test for age (p=0.0277) suggests MRS (MRS = 0.19 [CI: 0.09 / 0.30] vs. 0.15 [CI: 0.11 / 0.18) differs. Those with dependent adults would prioritize younger recipients more. The Wald test for dis2 (p=0.0000) suggests that MRS (MRS = 17.34 [CI: 11.86 / 22.83] vs. 14.77 [CI: 12.82 / 16.71]) may differ. This implies that those with moderate rather than severe disease affecting life expectancy are considered more of a priority by those with dependent adults.

5. Discussion.

Results from the baseline model (model 1) are broadly reassuring. They affirm the case for the revisions to UK transplant policy in 2006, which placed more emphasis upon factors such as prioritizing long waiters, and younger adults, which this analysis shows are priority criteria which patients value. Closeness of HLA tissue matching however remains an important consideration because the attribute is highly significant. It is somewhat surprising that patients do not value prioritizing those with no rather than moderate diseases affecting life expectancy significantly, but this could be due to the fact that many respondents may themselves have moderate diseases affecting life expectancy (so this finding ought to be treated with caution). The fact that respondents do value prioritizing those with moderate rather than severe diseases affecting life expectancy, makes sense, and is in line with UK transplant policy. It is somewhat surprising that respondents would prioritize those with

moderate illnesses affecting quality of life rather than no disease (other then kidney disease) affecting quality of life. However this may be because many respondents themselves have impaired quality of life and don't want to discriminate against people like themselves. Moreover, some respondents may believe that there is a case for helping those with moderate diseases affecting quality of life, in preference to those with no disease other than kidney disease. This is because dialysis itself adversely affects quality of life, and therefore a case could be made for reducing the suffering of those with other illnesses affecting their quality of life as a first priority, so their quality of life is not doubly impaired (by having to have ongoing dialysis). The clear preference for prioritizing those with moderate rather than severe diseases affecting quality of life makes sense, and is in line with UK transplant policy. It is interesting that patients seem to value prioritizing those with child or adult dependents quite a lot, and that this is not a criterion which currently explicitly figures in UK transplant policy, but perhaps ought to do.

There is some evidence that those who are successfully transplanted have different priorities (model 2). This is apparent because dummy variables are jointly significant according to the likelihood ratio test, and 3 / 8 dummy variables are significant. Moreover, Wald tests suggest that MRS differs for 3 variables. However none of the differences in MRS are dramatic. There is limited evidence that those with a failed transplant (model 3) have preferences which differ as the dummy variables are jointly insignificant. MRS though differs for 3 variables and quite markedly with respect to valuing a differences between having moderate rather than severe diseases affecting quality of life, which is valued less by the failed transplant group.

There is little evidence that those awaiting a transplant have different preferences (model 4), because the dummy variables are not jointly significant according to the LR test, and none of the dummy variables are significant (so variable impacts do not differ significantly). The Wald

tests however suggest that MRS may differ significantly for 5 variables, but MRS does not differ very much for any of them.

There is evidence that age affects preferences (model 5). The dummy variables for those in the lowest and highest quartiles are jointly significant according to a likelihood ratio test. Also 4 / 16 interaction dummy variables are significant, and MRS varies from the quartile range for 11 / 16 of the dummy variables. The most noteworthy finding is that those in the lowest quartile of the age distribution may value prioritizing younger respondents less than those in the inter-quartile range, whilst those in the highest quartile of the age distribution, value prioritizing the young more. This is a surprising result which we would not expect if respondents operated out of enlightened self-interest. The importance of respondent age as a determinant of preferences is confirmed by model 6 which compares preferences for those with above and below average age. The dummy variables are jointly significant according to the likelihood ratio test, and 2 / 8 interaction dummy variables proved significant, and MRS was always significantly different for every variable. Most worthy of note is the fact that respondents with above average age place a slightly higher valuation upon prioritizing transplants to younger recipients.

The evidence suggests (model 7) that the dummy variables for respondents with dependent children are not jointly significant according to a likelihood ratio test, and only the waiting time dummy variable is individually significant, although MRS differs significantly for 5 / 7 variables. The finding most worthy of note is that the MRS figures suggest that respondents with dependent children value those with dependents (either adults or children) less than those who don't have dependent children, a finding which is difficult to understand. Finally there is evidence (model 8) that the dummy variables for respondents with dependent adults are not jointly significant according to the likelihood ratio test. Moreover, none of the individual

dummy variables are significant, although 4 / 7 MRS scores are different comparing the 2 groups. The most important issue to address is whether those with dependent adults, value prioritizing those with either dependent adults or children more. The evidence suggests they do.

6. Conclusions.

The results remain broadly supportive of the reforms made in 2006 to UK renal transplantation policy, towards affording greater priority to those who had been on the transplant list longer, and to the young. However we have argued (chapter 5), that another consideration (whether recipients have child or adult dependents) ought to be given more explicit consideration when prioritizing people for kidney transplants.

Our findings do not seem to suggest that transplant status (whether you have had a successful transplant, failed transplant, or waiting a transplant) affects preferences in a manner that would make much of a difference to the general thrust of our conclusions. Therefore our findings would not be very sensitive to any imbalance in the number of successfully transplanted, failed transplant patients, or those awaiting a transplant, that might be present in our sample.

There is evidence that the age of patients affects their preferences. However, the fact that older respondents would prioritize younger recipients runs counter to theories of enlightened self-interested behaviour. Similarly patients with dependent children value prioritizing those with dependents less. However, in keeping with theories of enlightened self-interest, those patients with dependent adults value prioritizing those with dependents more highly.

7. References.

Byrne, C., D. Ford, et al. (2008). "ESRD incident rates in 2008: national and centre-specific analyses." **Chapter 3, UK Renal Registry Report [http://www.renalreg.org]**.

Byrne, C., R. Steenkamp, et al. (2008). "ESRD prevalent rates in 2008 national and centre-specific analyses." **Chapter 4, UK Renal Registry report [http://www.renalreg.org]**.

Wooldridge, J. M. (2002). "Economic analysis of cross section and panel data." MIT press.

HEALTH CARE PROVISION FOR SUSPECTED

DEEP VEIN THROMBOSIS

A SURVEY OF YOUR PREFERENCES

Undertaken by:

Centre for Health Services Studies,
Warwick Business School,
University of Warwick,
Coventry,
UK.

THE UNIVERSITY OF WARWICK

Undertaken for:

Leicester City West Primary Care Trust,
Ground Floor,
Mansion House,
41, Guildhall Lane,
Leicester,
LE1 5FR.

Leicester City West NHS
Primary Care Trust

What is this questionnaire about?

In this questionnaire we are trying to establish how health service provision for patients who may have Deep Vein Thrombosis (DVT: a blood clot in the leg) can be improved. There are different ways of providing a service for patients with suspected DVT. It is important therefore for us to ensure that any service we provide meets patient expectations as closely as possible.

In order to do this we have developed this questionnaire. The questionnaire does **not** attempt to evaluate the service you actually had, but instead is trying to find out your preferences for a number of different hypothetical (illustrative) health care scenarios for treatment of patients with DVT. Please begin by familiarising yourself with the various possible alternatives by **reading the box below.** Please bear in mind that all other factors about the service are **identical.**

Description of alternatives.

A) Continuity of staff.

Different ways of providing DVT services might affect whether a patient is likely to receive continuity in nursing provision. There could be:

1) Much continuity (same nurse for INR (blood) testing and counseling throughout).
2) Some continuity (mainly same nurse throughout for INR testing and counseling).
3) Lack of continuity (different nurses throughout for INR testing and counseling).

B) Number of visits to hospital.

Different ways of providing DVT services might influence / affect the number of trips patients need to make to hospital for routine INR (blood) tests, whilst they are treated for suspected or actual DVT. This may involve:

1) 0 visits to hospital
2) 4 visits to hospital
3) 8 visits to hospital
4) 12 visits to hospital

C) Hours during which routine dedicated DVT nurse provision is available.

The hours of routine dedicated DVT nurse provision is important because if such provision is not available, patients may instead be initially admitted to hospital. The hours of available DVT provision could be either:

1) 5 hours per day
2) 9 hours 30 minutes per day
3) 14 hours 30 minutes per day

D) Speed of imaging diagnosis.

The speed with which you receive an imaging diagnosis from a Venogram (x-ray with injection), or its alternative an Ultrasound scan, may impact upon the speed with which you

receive a definitive diagnosis of either having DVT, or not having DVT. About 66% of patients with suspected DVT in fact do not have DVT, and either a Venogram or Ultrasound scan will show this. Such patients can stop taking heparin, and cease having INR tests immediately after receiving either the Venogram or Ultrasound result. They can also be more quickly directed to more appropriate treatment (if required). The speed of imaging may be as follows:

1) Imaging results 24 hours after being suspected of having a DVT
2) Imaging results 3 days after being suspected of having a DVT
3) Imaging results 5 days after being suspected of having a DVT

E) Loss of income because of your illness.

There is absolutely no question of you being charged for your care. However we want you to imagine that your DVT provision may or may not involve you experiencing some loss of income, perhaps because of some time spent away from work. It could be that you will have:

1) No income lost
2) £50 of income lost
3) £150 of income lost
4) £250 of income lost

INSTRUCTIONS FOR COMPLETION OF THE QUESTIONNAIRE

- We now want you to choose between different patterns of service provision. Please remember that the questions are hypothetical, thus even if the provision you are receiving does not involve for example any income loss, or visits to hospital, we may ask you to consider DVT provision which does involve income loss or visits to hospital.

- It follows from this that we are not asking you to tick the pattern of care nearest to the actual pattern of care you received, we are interested in your preferences for these hypothetical scenarios.

- Everything else about the pattern of care apart from the stated differences is identical.

- Please answer every question remembering that there are no right or wrong answers, it is finding out what your personal preferences are that matters.

NOW PLEASE READ DESCRIPTIONS OF OPTION A AND B AND COMPLETE THE FOLLOWING 15 CHOICES:

1	DVT provision A	DVT provision B
Continuity of staff	Some	Much
Number of visits to hospital	0	12
Hours during which routine dedicated DVT nurse provision is available	9 hours 30 minutes	9 hours 30 minutes
Speed of imaging diagnosis	24 hours	24 hours
Loss of income	£50	£50

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

2	DVT provision A	DVT provision B
Continuity of staff	Lack	Lack
Number of visits to hospital	4	8
Hours during which routine dedicated DVT nurse provision is available	9 hours 30 minutes	14 hours 30 minutes
Speed of imaging diagnosis	24 hours	5 days
Loss of income	None	£50

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

3	DVT provision A	DVT provision B
Continuity of staff	Some	Lack
Number of visits to hospital	8	4
Hours during which routine dedicated DVT nurse provision is available	5 hours	5 hours
Speed of imaging diagnosis	24 hours	3 days
Loss of income	£150	£50

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

4	DVT provision A	DVT provision B
Continuity of staff	Lack	Much
Number of visits to hospital	12	12
Hours during which routine dedicated DVT nurse provision is available	5 hours	9 hours 30 minutes
Speed of imaging diagnosis	24 hours	24 hours
Loss of income	£250	£50

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

5	DVT provision A	DVT provision B
Continuity of staff	Much	Lack
Number of visits to hospital	0	0
Hours during which routine dedicated DVT nurse provision is available	5 hours	14 hours 30 minutes
Speed of imaging diagnosis	5 days	24 hours
Loss of income	None	£250

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

6	DVT provision A	DVT provision B
Continuity of staff	Lack	Some
Number of visits to hospital	8	12
Hours during which routine dedicated DVT nurse provision is available	9 hours 30 minutes	14 hours 30 minutes
Speed of imaging diagnosis	24 hours	3 days
Loss of income	None	None

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

7	DVT provision A	DVT provision B
Continuity of staff	Lack	Much
Number of visits to hospital	12	4
Hours during which routine dedicated DVT nurse provision is available	9 hours 30 minutes	14 hours 30 minutes
Speed of imaging diagnosis	5 days	24 hours
Loss of income	£150	£150

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

8	DVT provision A	DVT provision B
Continuity of staff	Some	Lack
Number of visits to hospital	4	0
Hours during which routine dedicated DVT nurse provision is available	9 hours 30 minutes	9 hours 30 minutes
Speed of imaging diagnosis	5 days	3 days
Loss of income	£250	£150

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

9	DVT provision A	DVT provision B
Continuity of staff	Some	Much
Number of visits to hospital	12	8
Hours during which routine dedicated DVT nurse provision is available	14 hours 30 minutes	9 hours 30 minutes
Speed of imaging diagnosis	3 days	3 days
Loss of income	None	£250

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

10	DVT provision A	DVT provision B
Continuity of staff	Much	Some
Number of visits to hospital	0 visits	8 visits
Hours during which routine dedicated DVT nurse provision is available	9 hours 30 minutes	5 hours
Speed of imaging diagnosis	24 hours	3 days
Loss of income	None	None

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

11	DVT provision A	DVT provision B
Continuity of staff	Lack	Some
Number of visits to hospital	8	8
Hours during which routine dedicated DVT nurse provision is available	14 hours 30 minutes	5 hours
Speed of imaging diagnosis	5 days	24 hours
Loss of income	£50	£150

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

12	DVT provision A	DVT provision B
Continuity of staff	Some	Some
Number of visits to hospital	12	4
Hours during which routine dedicated DVT nurse provision is available	14 hours 30 minutes	9 hours 30 minutes
Speed of imaging diagnosis	3 days	5 days
Loss of income	None	£250

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

13	DVT provision A	DVT provision B
Continuity of staff	Much	Much
Number of visits to hospital	0	12
Hours during which routine dedicated DVT nurse provision is available	5 hours	9 hours 30 minutes
Speed of imaging diagnosis	5 days	24 hours
Loss of income	None	£50

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

14	DVT provision A	DVT provision B
Continuity of staff	Lack	Much
Number of visits to hospital	12	4
Hours during which routine dedicated DVT nurse provision is available	5 hours	14 hours 30 minutes
Speed of imaging diagnosis	24 hours	24 hours
Loss of income	£250	£150

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

15	DVT provision A	DVT provision B
Continuity of staff	Lack	Some
Number of visits to hospital	4	4
Hours during which routine dedicated DVT nurse provision is available	5 hours	9 hours 30 minutes
Speed of imaging diagnosis	3 days	5 days
Loss of income	£50	£250

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

About you and your circumstances.

Can you currently drive? Yes ❑ No ❑

Do you have access to a car to travel to hospital? Yes ❑ No ❑ Sometimes ❑

When you last visited hospital in relation to your suspected DVT how did you travel to the hospital? *(Please tick the box that applies):*

1) Ambulance ❑
2) Car ❑
3) Taxi ❑
4) Bus ❑
5) Walked ❑
6) Other *(please describe_____).* ❑

How long did it take you to travel to this hospital?: _____ minutes

If you traveled by bus, taxi, or train, what is the fare both ways?: £_____.

Which hospital did you normally attend for your Deep Vein Thrombosis provision? *(Please tick 1 box only):*

Leicester Royal Infirmary	❑	Leicester General Hospital	❑
Glenfield General Hospital	❑	Other *(please describe):* _____	❑

Approximately how many miles from your home is the hospital?:_____miles.

When you had a suspected Deep Vein Thrombosis did you become an in-patient?

Yes ❑ No ❑

If 'Yes' for how many days?: _____days.

Gender:

(Please tick 1 box only): Male ❑ Female ❑

Age: ____ years

Do you have any children under 18 years?

(Please tick 1 box only):

Yes ❑ No ❑

If 'Yes' how many live in your household?: _____.

How many adults are there in the household?: _____.

Which of the following ethnic groups do you consider that you belong to?
(Please tick 1 box only):

White	❑
White and other	❑
Black – Caribbean	❑
Black – African	❑
Black – Other	❑
Indian	❑
Pakistani	❑
Bangladeshi	❑
Asian – Other	❑
Chinese	❑
Any other ethnic group (Please describe below):	❑

_____

Income:

Could you please estimate the annual income of your household, before deducting tax, and National Insurance *(Please tick 1 box only):*

Less than £5000	❑
£5,000 –£11,999	❑
£12,000-£19,999	❑
£20,000-£29,999	❑
£30,000-£44,999	❑
£45,000-£59,999	❑
£60,000+	❑

Work:

Are you currently in paid employment? *(Please tick 1 box only):*

Yes (full time) ❑ Yes (part-time) ❑ No ❑

Qualifications:

What is the highest level of education you have completed *(Please tick 1 box only):*

Secondary School	❑
Vocational / Trade / College Qualification	❑
'A' level / 'AS' levels	❑
Degree level qualification(s)	❑
Other *(please describe*_____*)*.	❑

Would you be prepared to fill in another similar questionnaire in the future to help us to validate our methodology? (Please note, the researchers will not be able to identify any patients from the questionnaire. If you agree to participate in future research by the NHS trust, at no stage will your identity or identifying details be given to anyone who is not involved in your care).

Yes ❑ No ❑

THANK YOU FOR COMPLETING THIS QUESTIONNAIRE
PLEASE PUT IT IN THE FREEPOST ENVELOPE PROVIDED, AND PUT IT IN THE POST.

Appendix C

REF:

HEALTH CARE FOR WOMEN WITH PERIOD PROBLEMS

A SURVEY OF WHAT YOU PREFER

By:

Dr Sophia Julian,
Department of Obstetrics and Gynaecology, and the University of Leicester Medical School,
Robert Kilpatrick Building,
Leicester Royal Infirmary,
Leicester, LE2 7LX.

Mr Michael Clark
Senior Research Fellow,
Centre for Health Services Studies,
Warwick Business School,
University of Warwick,
Coventry,
CV4 7AL.

This is a questionnaire to find out what women would like when they get help for their period problems.

We hope you will help us with this important research by filling in this survey and returning it in the FREEPOST envelope provided.

Instructions:

- We want you to choose between different types of health care. The questions are made up examples of different types of health care we could offer to women. This means that we will be asking you to think about things that might not have happened to you.
- We are asking you how you would have like to get health care for a period problem not what health care you have had in the past.
- There are no right or wrong answers, we just want to know what you prefer.
- Your choices are important to us. We can only make things better if we find out what people want. Filling in this survey will not affect your health care, but you will be helping us to help other women in the future.
- This survey is confidential. None of the doctors or nurses who have treated you will find out your answers.

What you need to do:

- In section A there are 8 parts. Each part has two different examples of the kind of option which women with period problems could be offered.
- For each part we want you to tell us the option you would prefer. Put a tick underneath the list for the option you would prefer i.e. put a tick either under option A or B.
- In section B we ask a few brief questions.
- The questionnaire usually takes about 20 minutes to fill in.
- Many people find filling in forms difficult. If you would like some help filling in this survey please phone me. I will be very happy to phone you back and go through it with you, or answer any questions you might have.
- Thank you in advance for your help.

Dr. Sophia Julian

Department of Obstetrics and Gynaecology

Tel (0116) 252 5883 Email slj2@le.ac.uk

Section A

In Section A, these are the things that we want you to think about.

The thing that may vary:	Could be:
How long you have to wait for test results	1 day 2 days 2 weeks 4 weeks
Cost to you (i.e. perhaps because of absence from work or travel costs - <u>Please assume you would loose this amount of money even if you would not).</u>	None (no money lost) £25 £75 £125
The type of doctor you see	GP or Consultant
The sex of the doctor you see	Male or Female
Time waiting for an appointment to see the doctor (either the GP or the consultant)	1 day 4 days 6 weeks 12 weeks
How often you get to see the same doctor	None of the time Half of the time All of the time

- **Note there is no question of you being charged for health care, but <u>we want you to pretend that you would lose the amount of money shown even if you would not.</u>**
- **Just a few of the possible combinations are included, these are chosen by a computer programme.**
- **Remember, everything else apart from the things on the list, like the receptionists, nurses and waiting area etc, is the same for option A and B.**
- **Remember there are no right or wrong answers.**

Please read the whole descriptions of option A and B below and for each of the parts choose A or B.

Part 1	Option A	Option B
How long you have to wait for test results	1 day	2 weeks
Cost to you	None	None
The type of doctor you see	Consultant	GP
The sex of the doctor you see	Male	Male
Time waiting for an appointment to see the doctor	12 weeks	6 weeks
How often you get to see the same doctor	Half of the time	All of the time

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 2	Option A	Option B
How long you have to wait for test results	1 day	1 day
Cost to you	None	£25
The type of doctor you see	Consultant	GP
The sex of the doctor you see	Male	Female
Time waiting for an appointment to see the doctor	12 weeks	1 day
How often you get to see the same doctor	Half of the time	All of the time

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 3	Option A	Option B
How long you have to wait for test results	1 day	4 weeks
Cost to you	None	£125
The type of doctor you see	Consultant	Consultant
The sex of the doctor you see	Male	Female
Time waiting for an appointment to see the doctor	12 weeks	12 weeks
How often you get to see the same doctor	Half of the time	All of the time

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 4	Option A	Option B
How long you have to wait for test results	1 day	1 day
Cost to you	None	None
The type of doctor you see	Consultant	Consultant
The sex of the doctor you see	Male	Male
Time waiting for an appointment to see the doctor	12 weeks	6 weeks
How often you get to see the same doctor	Half of the time	Half of the time

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 5	Option A	Option B
How long you have to wait for test results	1 day	2 weeks
Cost to you	None	£25
The type of doctor you see	Consultant	Consultant
The sex of the doctor you see	Male	Female
Time waiting for an appointment to see the doctor	12 weeks	4 days
How often you get to see the same doctor	Half of the time	Half of the time

Which option would you choose? Choose A ❑ Choose B ❑
(tick 1 box only)

Part 6	Option A	Option B
How long you have to wait for test results	1 day	4 weeks
Cost to you	None	£25
The type of doctor you see	Consultant	Consultant
The sex of the doctor you see	Male	Male
Time waiting for an appointment to see the doctor	12 weeks	6 weeks
How often you get to see the same doctor	Half of the time	None of the time

Which option would you choose? Choose A ❑ Choose B ❑
(tick 1 box only)

Part 7	Option A	Option B
How long you have to wait for test results	1 day	4 weeks
Cost to you	None	None
The type of doctor you see	Consultant	GP
The sex of the doctor you see	Male	Female
Time waiting for an appointment to see the doctor	12 weeks	4 days
How often you get to see the same doctor	Half of the time	Half of the time

Which option would you choose? Choose A ❑ Choose B ❑
(tick 1 box only)

Part 8	Option A	Option B
How long you have to wait for test results	1 day	2 days
Cost to you	None	None
The type of doctor you see	Consultant	Consultant
The sex of the doctor you see	Male	Female
Time waiting for an appointment to see the doctor	12 weeks	1 day
How often you get to see the same doctor	Half of the time	None of the time

Which option would you choose? Choose A ❑ Choose B ❑
(tick 1 box only)

Section B

About Filling in the Questionnaire

When you answered this questionnaire did **one** thing matter to you more than anything else?

Yes ❑
No ❑

If Yes, which **one** mattered most? **If No,** go to next box.

	Tick one box only
Differences in the type of doctor you see (GP or Consultant)	❑
Differences in the sex of the doctor you see (Male or Female)	❑
Differences in time spent waiting for an appointment to see the doctor	❑
Differences in how often I would get to see the same doctor	❑
Differences in how long I have to wait for results	❑
Differences in 'cost to you'	❑

		Tick one box only
If Yes, did you ignore the other things on the list?	Yes	❑
	No	❑
	Sometimes	❑

Did differences in the amount of 'cost to you' for options A and B, influence your choices?

Yes ❑
No ❑
Sometimes ❑

Do you have a preference for either having a male or female doctor about a period problem?

Strong preference for Male	Some preference for male	No preference	Some preference for female	Strong preference for female
❑	❑	❑	❑	❑

If you have a preference, can your GP surgery **meet this preference?**

Yes ❑
No ❑
Sometimes ❑
Don't know ❑

How easy is it for you to get to:

Your GP's surgery?		**Leicester Royal Infirmary?**	
Easy access	❑	Easy access	❑
Reasonable access	❑	Reasonable access	❑
Difficult access	❑	Difficult access	❑

About You and Your Circumstances: To make sure we have women from all different backgrounds in the study.

Which ethnic group do you consider you belong to?

Tick one box only

White ❑
Includes British, Irish, and any other white background

Mixed ❑
Includes white and black Caribbean, white and black African, white and Asian; any mixed background

Asian or Asian British ❑
Includes Indian; Pakistani; Bangladeshi; any other Asian background

Black or Black British ❑
Includes Caribbean; African; or other black background

Other ethnic group ❑
Includes Chinese or any other ethnic group – please describe below
……..

How old are you? **……** Years

Overall, what has happened to your **household** income **(after deducting income tax and national insurance)** in the last year?

Tick one box only

Stayed about the same ❑
Gone down ❑
Gone up ❑

The average annual household **income** before tax **in Leicestershire is about £25,000. Is your household income:**

Tick one box only

Above average ❑
About average ❑
Below average ❑

About Your Health Care

When you have been seen by a doctor in relation to a period problem has it tended to be:

Mainly a GP ❑
About the same likelihood of seeing a GP or Consultant ❑
Mainly a Consultant ❑

How long on average do you have to wait to see a GP about a period problem?

…… **Days**

How long on average do you have to wait to see a hospital doctor about a period problem?

…… **Days**
…… Weeks

How long on average have you had to wait for results of tests relating to period problems?

…… **Days**
…… Weeks

How often do you usually see the same doctor in relation to period problems?

None of the time ❑
Between none and half of the time ❑
Half of the time ❑
Between half and all of the time ❑
All of the time ❑

Would you normally see a male or a female GP about your period problems?

Always male ❑
Usually male ❑
About the same likelihood of seeing a male or female ❑
Usually female ❑
Always female ❑

Thank you for completing this questionnaire.

Please post it in the FREEPOST envelope provided.

REF:

HEALTH CARE FOR WOMEN WITH PERIOD PROBLEMS

A SURVEY OF WHAT YOU PREFER

By:

Dr Sophia Julian,
Department of Obstetrics and Gynaecology, and the University of Leicester Medical School,
Robert Kilpatrick Building,
Leicester Royal Infirmary,
Leicester, LE2 7LX.

Mr Michael Clark
Senior Research Fellow,
Centre for Health Services Studies,
Warwick Business School,
University of Warwick,
Coventry,
CV4 7AL.

This is a questionnaire to find out what women would like when they get help for their period problems.

We hope you will help us with this important research by filling in this survey and returning it in the FREEPOST envelope provided.

Instructions:

- We want you to choose between different types of health care. The questions are made up examples of different types of health care we could offer to women. This means that we will be asking you to think about things that might not have happened to you.
- We are asking you how you would have like to get health care for a period problem not what health care you have had in the past.
- There are no right or wrong answers, we just want to know what you prefer.
- Your choices are important to us. We can only make things better if we find out what people want. Filling in this survey will not affect your health care, but you will be helping us to help other women in the future.
- This survey is confidential. None of the doctors or nurses who have treated you will find out your answers.

What you need to do:

- In section A there are 8 parts. Each part has two different examples of the kind of option which women with period problems could be offered.
- For each part we want you to tell us the option you would prefer. Put a tick underneath the list for the option you would prefer i.e. put a tick either under option A or B.
- In section B we ask a few brief questions.
- The questionnaire usually takes about 20 minutes to fill in.
- Many people find filling in forms difficult. If you would like some help filling in this survey please phone me. I will be very happy to phone you back and go through it with you, or answer any questions you might have.
- Thank you in advance for your help.

Dr. Sophia Julian

Department of Obstetrics and Gynaecology

Tel (0116) 252 5883 Email slj2@le.ac.uk

Section A

In Section A, these are the things that we want you to think about.

The thing that may vary:	Could be:
How long you have to wait for test results	1 day 2 days 2 weeks 4 weeks
Cost to you (i.e. perhaps because of absence from work or travel costs - <u>Please assume you would loose this amount of money even if you would not).</u>	None (no money lost) £25 £75 £125
The type of doctor you see	GP or Consultant
The sex of the doctor you see	Male or Female
Time waiting for an appointment to see the doctor (either the GP or the consultant)	1 day 4 days 6 weeks 12 weeks
How often you get to see the same doctor	None of the time Half of the time All of the time

- **Note there is no question of you being charged for health care, but <u>we want you to pretend that you would lose the amount of money shown even if you would not.</u>**
- **Just a few of the possible combinations are included, these are chosen by a computer programme.**
- **Remember, everything else apart from the things on the list, like the receptionists, nurses and waiting area etc, is the same for option A and B.**
- **Remember there are no right or wrong answers.**

Please read the whole descriptions of option A and B below and for each of the parts choose A or B.

Part 1	Option A	Option B
How long you have to wait for test results	1 day	2 weeks
Cost to you	None	None
The type of doctor you see	Consultant	Consultant
The sex of the doctor you see	Male	Male
Time waiting for an appointment to see the doctor	12 weeks	12 weeks
How often you get to see the same doctor	Half of the time	Half of the time

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 2	Option A	Option B
How long you have to wait for test results	1 day	1 day
Cost to you	None	£75
The type of doctor you see	Consultant	Consultant
The sex of the doctor you see	Male	Female
Time waiting for an appointment to see the doctor	12 weeks	6 weeks
How often you get to see the same doctor	Half of the time	Half of the time

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 3	Option A	Option B
How long you have to wait for test results	1 day	1 day
Cost to you	None	£125
The type of doctor you see	Consultant	GP
The sex of the doctor you see	Male	Male
Time waiting for an appointment to see the doctor	12 weeks	4 days
How often you get to see the same doctor	Half of the time	None of the time

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 4	Option A	Option B
How long you have to wait for test results	1 day	2 days
Cost to you	None	£75
The type of doctor you see	Consultant	Consultant
The sex of the doctor you see	Male	Male
Time waiting for an appointment to see the doctor	12 weeks	4 days
How often you get to see the same doctor	Half of the time	All of the time

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 5	Option A	Option B
How long you have to wait for test results	1 day	2 days
Cost to you	None	£125
The type of doctor you see	Consultant	GP
The sex of the doctor you see	Male	Female
Time waiting for an appointment to see the doctor	12 weeks	6 weeks
How often you get to see the same doctor	Half of the time	Half of the time

Which option would you choose? Choose A ❑ Choose B ❑
(tick 1 box only)

Part 6	Option A	Option B
How long you have to wait for test results	1 day	2 weeks
Cost to you	None	£75
The type of doctor you see	Consultant	GP
The sex of the doctor you see	Male	Female
Time waiting for an appointment to see the doctor	12 weeks	12 weeks
How often you get to see the same doctor	Half of the time	None of the time

Which option would you choose? Choose A ❑ Choose B ❑
(tick 1 box only)

Part 7	Option A	Option B
How long you have to wait for test results	1 day	4 weeks
Cost to you	None	£75
The type of doctor you see	Consultant	GP
The sex of the doctor you see	Male	Male
Time waiting for an appointment to see the doctor	12 weeks	1 day
How often you get to see the same doctor	Half of the time	Half of the time

Which option would you choose? Choose A ❑ Choose B ❑
(tick 1 box only)

Part 8	Option A	Option B
How long you have to wait for test results	1 day	2 days
Cost to you	None	£25
The type of doctor you see	Consultant	GP
The sex of the doctor you see	Male	Male
Time waiting for an appointment to see the doctor	12 weeks	12 weeks
How often you get to see the same doctor	Half of the time	Half of the time

Which option would you choose? Choose A ❑ Choose B ❑
(tick 1 box only)

Section B

About Filling in the Questionnaire

When you answered this questionnaire did **one** thing matter to you more than anything else?

Yes ❑
No ❑

If Yes, which **one** mattered most? **If No,** go to next box.

	Tick one box only
Differences in the type of doctor you see (GP or Consultant)	❑
Differences in the sex of the doctor you see (Male or Female)	❑
Differences in time spent waiting for an appointment to see the doctor	❑
Differences in how often I would get to see the same doctor	❑
Differences in how long I have to wait for results	❑
Differences in 'cost to you'	❑

		Tick one box only
If Yes, did you ignore the other things on the list?	Yes	❑
	No	❑
	Sometimes	❑

Did differences in the amount of 'cost to you' for options A and B, influence your choices?

Yes ❑
No ❑
Sometimes ❑

Do you have a preference for either having a male or female doctor about a period problem?

Strong preference for Male	Some preference for male	No preference	Some preference for female	Strong preference for female
❑	❑	❑	❑	❑

If you have a preference, can your GP surgery **meet this preference?**

Yes ❑
No ❑
Sometimes ❑
Don't know ❑

How easy is it for you to get to:

Your GP's surgery?		**Leicester Royal Infirmary?**	
Easy access	❑	Easy access	❑
Reasonable access	❑	Reasonable access	❑
Difficult access	❑	Difficult access	❑

About You and Your Circumstances: To make sure we have women from all different backgrounds in the study.

Which ethnic group do you consider you belong to?

Tick one box only

White ❑
Includes British, Irish, and any other white background

Mixed ❑
Includes white and black Caribbean, white and black African, white and Asian; any mixed background

Asian or Asian British ❑
Includes Indian; Pakistani; Bangladeshi; any other Asian background

Black or Black British ❑
Includes Caribbean; African; or other black background

Other ethnic group ❑
Includes Chinese or any other ethnic group – please describe below
……..

How old are you? **……** Years

Overall, what has happened to your **household** income **(after deducting income tax and national insurance)** in the last year?

Tick one box only

Stayed about the same ❑
Gone down ❑
Gone up ❑

The average annual household **income** before tax **in Leicestershire is about £25,000. Is your household income:**

Tick one box only

Above average ❑
About average ❑
Below average ❑

About Your Health Care

When you have been seen by a doctor in relation to a period problem has it tended to be:

Mainly a GP ❑
About the same likelihood of seeing a GP or Consultant ❑
Mainly a Consultant ❑

How long on average do you have to wait to see a GP about a period problem?

…… **Days**

How long on average do you have to wait to see a hospital doctor about a period problem?

…… **Days**
…… Weeks

How long on average have you had to wait for results of tests relating to period problems?

…… **Days**
…… Weeks

How often do you usually see the same doctor in relation to period problems?

None of the time ❑
Between none and half of the time ❑
Half of the time ❑
Between half and all of the time ❑
All of the time ❑

Would you normally see a male or a female GP about your period problems?

Always male ❑
Usually male ❑
About the same likelihood of seeing a male or female ❑
Usually female ❑
Always female ❑

Thank you for completing this questionnaire.

Please post it in the FREEPOST envelope provided.

Appendix D

REF:

GYNAECOLOGY HEALTH CARE

A SURVEY OF WHAT YOU

PREFER

By:

Dr Sophia Julian
Department of Obs & Gynae
Leicester-Warwick Medical School
Robert Kilpatrick Building
Leicester Royal Infirmary
Leicester
LE2 7LX

Mr Michael Clark
Senior Research Fellow
Centre for Health Services Studies,
Warwick Business School
University of Warwick
Coventry,
CV4 7AL

This is a questionnaire to find out what women would like when they need gynaecological healthcare.

We hope you will help us with this important research by filling in this survey.

Instructions:

- We want you to choose between different types of health care. The questions are made up examples of different types of health care we could offer to women. This means that we will be asking you to think about things that might not have happened to you.
- We are asking you how you would have like to get gynaecological care not what health care you have had in the past.
- There are no right or wrong answers, we just want to know what you prefer.
- Your choices are important to us. We can only make things better if we find out what people want. Filling in this survey will not affect your health care, but you will be helping us to help other women in the future.
- This survey is confidential. Do not put your name on it. None of the doctors or nurses who have treated you will find out your answers.

What you need to do:

- In section A there are 8 parts. Each part has two different examples of the kind of option which women requiring gynaecological care could be offered. The options might look the same at a glance, but they are all different.
- For each part we want you to tell us the option you would prefer. Put a tick underneath the list for the option you would prefer i.e. put a tick either under option A or B.
- In section B there are questions about you and your health care.
- The questionnaire usually takes about 20 minutes to fill in.
- If you have any questions, please see the receptionist.

Thank you in advance for your help

Section A

In Section A, these are the things that we want you to think about.

The thing that may vary	**Could be**
Amount lost (i.e. perhaps because of absence from work or travel costs - Please assume you would loose this amount of money even if you would not).	None (no money lost) £25 £75 £125
The type of doctor you see	GP or Consultant
The sex of the doctor you see	Male or Female
Time waiting for an appointment to see the doctor (either the GP or the consultant)	1 day 4 days 6 weeks 12 weeks
How often you get to see the same doctor	None of the time Half of the time All of the time
How long you have to wait for test results	1 day 2 days 2 weeks 4 weeks

- Note there is no question of you being charged for health care, but we want you to pretend that you would lose the amount of money shown even if you would not.
- Just a few of the possible combinations are included, these are chosen by a computer programme.
- Remember, everything else apart from the things on the list, like the receptionists, nurses and waiting area etc, is the same for option A and B.
- Remember there are no right or wrong answers.

Please read the whole descriptions of option A and B below and for each of the parts choose A or B.

Part 1	Option A	Option B
Amount lost	None	None
The type of doctor you see	Consultant	GP
The sex of the doctor you see	Male	Male
Time waiting for an appointment to see the doctor	12 weeks	6 weeks
How often you get to see the same doctor	Half of the time	All of the time
How long you have to wait for test results	1 day	2 weeks

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 2	Option A	Option B
Amount lost	None	£25
The type of doctor you see	Consultant	GP
The sex of the doctor you see	Male	Female
Time waiting for an appointment to see the doctor	12 weeks	1 day
How often you get to see the same doctor	Half of the time	All of the time
How long you have to wait for test results	1 day	1 day

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 3	Option A	Option B
Amount lost	None	£125
The type of doctor you see	Consultant	Consultant
The sex of the doctor you see	Male	Female
Time waiting for an appointment to see the doctor	12 weeks	12 weeks
How often you get to see the same doctor	Half of the time	All of the time
How long you have to wait for test results	1 day	4 weeks

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 4	Option A	Option B
Amount lost	None	None
The type of doctor you see	Consultant	Consultant
The sex of the doctor you see	Male	Male
Time waiting for an appointment to see the doctor	12 weeks	6 weeks
How often you get to see the same doctor	Half of the time	Half of the time
How long you have to wait for test results	1 day	1 day

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 5	Option A	Option B
Amount lost	None	£25
The type of doctor you see	Consultant	Consultant
The sex of the doctor you see	Male	Female
Time waiting for an appointment to see the doctor	12 weeks	4 days
How often you get to see the same doctor	Half of the time	Half of the time
How long you have to wait for test results	1 day	2 weeks

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 6	Option A	Option B
Amount lost	None	£25
The type of doctor you see	Consultant	Consultant
The sex of the doctor you see	Male	Male
Time waiting for an appointment to see the doctor	12 weeks	6 weeks
How often you get to see the same doctor	Half of the time	None of the time
How long you have to wait for test results	1 day	4 weeks

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 7	Option A	Option B
Amount lost	None	None
The type of doctor you see	Consultant	GP
The sex of the doctor you see	Male	Female
Time waiting for an appointment to see the doctor	12 weeks	4 days
How often you get to see the same doctor	Half of the time	Half of the time
How long you have to wait for test results	1 day	4 weeks

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 8	Option A	Option B
Amount lost	None	None
The type of doctor you see	Consultant	Consultant
The sex of the doctor you see	Male	Female
Time waiting for an appointment to see the doctor	12 weeks	1 day
How often you get to see the same doctor	Half of the time	None of the time
How long you have to wait for test results	1 day	2 days

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Section B

About Filling in the Questionnaire

When you answered this questionnaire did **one** thing matter to you more than anything else?

- ❑ Yes
- ❑ No

If Yes, which **one** mattered most? **If No,** go to the next box.

- ❑ Differences in the type of doctor you see (GP or Consultant)
- ❑ Differences in how long I have to wait for results
- ❑ Differences in the sex of the doctor you see (Male or Female)
- ❑ Differences in how often I would get to see the same doctor
- ❑ Differences in time spent waiting for an appointment to see the doctor
- ❑ Differences in 'amount lost'

If Yes, did you ignore the other things on the list?

- ❑ Yes
- ❑ No
- ❑ Sometimes

Did differences in the amount of 'amount lost' for options A and B influence your choices?

- ❑ Yes
- ❑ No
- ❑ Sometimes

Do you have a preference for either seeing a male or female doctor about a gynaecological problem

- ❑ Strong preference for male
- ❑ Some preference for male
- ❑ No preference
- ❑ Some preference for female
- ❑ Strong preference for female

If you have a preference, can your **GP surgery** meet this preference?

- ❑ Yes
- ❑ No
- ❑ Sometimes
- ❑ Don't know

About You and Your Circumstances: To make sure we have women from all different backgrounds in the study.

How easy is it for you to get to:

Your GP's surgery?

- ❑ Easy access
- ❑ Reasonable access
- ❑ Difficult access

Leicester Royal Infirmary?

- ❑ Easy access
- ❑ Reasonable access
- ❑ Difficult access

Which ethnic group do you consider you belong to?

- ❑ **White**
  Includes British, Irish, and any other white background
- ❑ **Mixed**
  Includes white and black Caribbean, white and black African, white and Asian; any mixed background
- ❑ **Asian or Asian British**
  Includes Indian; Pakistani; Bangladeshi; any other Asian background
- ❑ **Black or Black British**
  Includes Caribbean; African; or other black background
- ❑ **Other ethnic group**
  Includes Chinese or any other ethnic group – please describe below
  ……

How old are you?

…… Years

Overall, what has happened to your **household** income **(after deducting income tax and national insurance)** in the last year?

- ❑ Stayed about the same
- ❑ Gone down
- ❑ Gone up

The average annual household income before tax in Leicestershire is about £25,000.
Is your household income:

- ❑ Above average
- ❑ About average
- ❑ Below average

About Your Health Care

When you have been seen by a doctor in relation to a gynaecological problem to be:

- ❑ Mainly a GP
- ❑ About the same likelihood of seeing a GP or Consultant
- ❑ Mainly a Consultant

How long on average do you have to wait to see a GP about a gynaecological problem?

…… Days

How long on average do you have to wait to see a hospital doctor about a gynaecological problem?

…… Days
…… Weeks

How long on average have you had to wait for results of tests relating to gynaecological problem?

…… Days
…… Weeks

How often do you usually see the same doctor in relation to gynaecological problem?

- ❑ None of the time
- ❑ Between none and half of the time
- ❑ Half of the time
- ❑ Between half and all of the time
- ❑ All of the time

Would you normally see a male or a female GP about your gynaecological problem?

- ❑ Always male
- ❑ Usually male
- ❑ About the same likelihood of seeing a male or female
- ❑ Usually female
- ❑ Always female

Thank you

Please put it in the envelope and then in the box at reception

REF:

GYNAECOLOGY HEALTH CARE

A SURVEY OF WHAT YOU

PREFER

By:

Dr Sophia Julian
Department of Obs & Gynae
Leicester-Warwick Medical School
Robert Kilpatrick Building
Leicester Royal Infirmary
Leicester
LE2 7LX

Mr Michael Clark
Senior Research Fellow
Centre for Health Services Studies,
Warwick Business School
University of Warwick
Coventry,
CV4 7AL

This is a questionnaire to find out what women would like when they need gynaecological healthcare.

We hope you will help us with this important research by filling in this survey.

Instructions:

- We want you to choose between different types of health care. The questions are made up examples of different types of health care we could offer to women. This means that we will be asking you to think about things that might not have happened to you.
- We are asking you how you would have like to get gynaecological care not what health care you have had in the past.
- There are no right or wrong answers, we just want to know what you prefer.
- Your choices are important to us. We can only make things better if we find out what people want. Filling in this survey will not affect your health care, but you will be helping us to help other women in the future.
- This survey is confidential. Do not put your name on it. None of the doctors or nurses who have treated you will find out your answers.

What you need to do:

- In section A there are 8 parts. Each part has different examples of the kind of option which women requiring gynaecological care could be offered. The options might look the same at a glance, but they are all different.
- For each part we want you to tell us the option you would prefer. Put a tick underneath the list for the option you would prefer i.e. put a tick either under option A or B.
- In section B there are questions about you and your health care.
- The questionnaire usually takes about 20 minutes to fill in.
- If you have any questions, please see the receptionist.

Thank you in advance for your help

Section A

In Section A, these are the things that we want you to think about.

The thing that may vary	**Could be**
Willingness to pay (i.e. perhaps because of absence from work or travel costs - Please assume you would loose this amount of money even if you would not).	None (no money lost) £25 £75 £125
The type of doctor you see	GP or Consultant
The sex of the doctor you see	Male or Female
Time waiting for an appointment to see the doctor (either the GP or the consultant)	1 day 4 days 6 weeks 12 weeks
How often you get to see the same doctor	None of the time Half of the time All of the time
How long you have to wait for test results	1 day 2 days 2 weeks 4 weeks

- Note there is no question of you being charged for health care, but we want you to pretend that you would lose the amount of money shown even if you would not.
- Just a few of the possible combinations are included, these are chosen by a computer programme.
- Remember, everything else apart from the things on the list, like the receptionists, nurses and waiting area etc, is the same for option A and B.
- Remember there are no right or wrong answers.

Please read the whole descriptions of option A and B below and for each of the parts choose A or B.

Part 1	Option A	Option B
Willingness to pay	None	None
The type of doctor you see	Consultant	Consultant
The sex of the doctor you see	Male	Male
Time waiting for an appointment to see the doctor	12 weeks	12 weeks
How often you get to see the same doctor	Half of the time	Half of the time
How long you have to wait for test results	1 day	2 weeks

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 2	Option A	Option B
Willingness to pay	None	£75
The type of doctor you see	Consultant	Consultant
The sex of the doctor you see	Male	Female
Time waiting for an appointment to see the doctor	12 weeks	6 weeks
How often you get to see the same doctor	Half of the time	Half of the time
How long you have to wait for test results	1 day	1 day

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 3	Option A	Option B
Willingness to pay	None	£125
The type of doctor you see	Consultant	GP
The sex of the doctor you see	Male	Male
Time waiting for an appointment to see the doctor	12 weeks	4 days
How often you get to see the same doctor	Half of the time	None of the time
How long you have to wait for test results	1 day	1 day

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 4	Option A	Option B
Willingness to pay	None	£75
The type of doctor you see	Consultant	Consultant
The sex of the doctor you see	Male	Male
Time waiting for an appointment to see the doctor	12 weeks	4 days
How often you get to see the same doctor	Half of the time	All of the time
How long you have to wait for test results	1 day	2 days

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 5	Option A	Option B
Willingness to pay	None	£125
The type of doctor you see	Consultant	GP
The sex of the doctor you see	Male	Female
Time waiting for an appointment to see the doctor	12 weeks	6 weeks
How often you get to see the same doctor	Half of the time	Half of the time
How long you have to wait for test results	1 day	2 days

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 6	Option A	Option B
Willingness to pay	None	£75
The type of doctor you see	Consultant	GP
The sex of the doctor you see	Male	Female
Time waiting for an appointment to see the doctor	12 weeks	12 weeks
How often you get to see the same doctor	Half of the time	None of the time
How long you have to wait for test results	1 day	2 weeks

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 7	Option A	Option B
Willingness to pay	None	£75
The type of doctor you see	Consultant	GP
The sex of the doctor you see	Male	Male
Time waiting for an appointment to see the doctor	12 weeks	1 day
How often you get to see the same doctor	Half of the time	Half of the time
How long you have to wait for test results	1 day	4 weeks

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Part 8	Option A	Option B
Willingness to pay	None	£25
The type of doctor you see	Consultant	GP
The sex of the doctor you see	Male	Male
Time waiting for an appointment to see the doctor	12 weeks	12 weeks
How often you get to see the same doctor	Half of the time	Half of the time
How long you have to wait for test results	1 day	2 days

Which option would you choose? (tick 1 box only) Choose A ❑ Choose B ❑

Section B

About Filling in the Questionnaire

When you answered this questionnaire did **<u>one</u>** thing matter to you more than anything else?

- ❑ Yes
- ❑ No

If Yes, which **<u>one</u>** mattered most? **If No,** go to the next box.

- ❑ Differences in the type of doctor you see (GP or Consultant)
- ❑ Differences in how long I have to wait for results
- ❑ Differences in the sex of the doctor you see (Male or Female)
- ❑ Differences in how often I would get to see the same doctor
- ❑ Differences in time spent waiting for an appointment to see the doctor
- ❑ Differences in 'Willingness to pay'

If Yes, did you ignore the other things on the list?

- ❑ Yes
- ❑ No
- ❑ Sometimes

Did differences in the amount of 'Willingness to pay' for options A and B influence your choices?

- ❑ Yes
- ❑ No
- ❑ Sometimes

Do you have a preference for either seeing a male or female doctor about a gynaecological problem

- ❑ Strong preference for male
- ❑ Some preference for male
- ❑ No preference
- ❑ Some preference for female
- ❑ Strong preference for female

If you have a preference, can your **GP surgery** meet this preference?

- ❑ Yes
- ❑ No
- ❑ Sometimes
- ❑ Don't know

About You and Your Circumstances: To make sure we have women from all different backgrounds in the study.

How easy is it for you to get to:

Your GP's surgery?

- ❑ Easy access
- ❑ Reasonable access
- ❑ Difficult access

Leicester Royal Infirmary?

- ❑ Easy access
- ❑ Reasonable access
- ❑ Difficult access

Which ethnic group do you consider you belong to?

- ❑ **White**
  Includes British, Irish, and any other white background
- ❑ **Mixed**
  Includes white and black Caribbean, white and black African, white and Asian; any mixed background
- ❑ **Asian or Asian British**
  Includes Indian; Pakistani; Bangladeshi; any other Asian background
- ❑ **Black or Black British**
  Includes Caribbean; African; or other black background
- ❑ **Other ethnic group**
  Includes Chinese or any other ethnic group – please describe below
  ……

How old are you?

…… Years

Overall, what has happened to your **household** income **(after deducting income tax and national insurance)** in the last year?

- ❑ Stayed about the same
- ❑ Gone down
- ❑ Gone up

The average annual household income before tax in Leicestershire is about £25,000.
Is your household income:

- ❑ Above average
- ❑ About average
- ❑ Below average

About Your Health Care

When you have been seen by a doctor in relation to a gynaecological problem to be:

- ❑ Mainly a GP
- ❑ About the same likelihood of seeing a GP or Consultant
- ❑ Mainly a Consultant

How long on average do you have to wait to see a GP about a gynaecological problem?

…… Days

How long on average do you have to wait to see a hospital doctor about a gynaecological problem?

…… Days
…… Weeks

How long on average have you had to wait for results of tests relating to gynaecological problem?

…… Days
…… Weeks

How often do you usually see the same doctor in relation to gynaecological problem?

- ❑ None of the time
- ❑ Between none and half of the time
- ❑ Half of the time
- ❑ Between half and all of the time
- ❑ All of the time

Would you normally see a male or a female GP about your gynaecological problem?

- ❑ Always male
- ❑ Usually male
- ❑ About the same likelihood of seeing a male or female
- ❑ Usually female
- ❑ Always female

Thank you

Please put it in the envelope and then in the box at reception

Appendix E

REF:

Who should be prioritized for Kidney Transplants in the UK?

A SURVEY OF YOUR PREFERENCES.

Undertaken by:

Warwick Medical School,
University of Warwick,
Coventry,
UK,
CV4 7AL.

Undertaken for:

University Hospitals of Coventry
and Warwickshire NHS Trust,
Walsgrave Hospital,
Clifford Bridge Road,
Coventry,
CV2 2DX

Version 4a: Patient version: 16/05/2006

What is this questionnaire about?:

In this questionnaire we are trying to establish how to prioritize patients for a renal transplant. It is important for us to ensure that the systems for kidney allocation take account of the views of patients, health professionals and others involved in the transplant process. This questionnaire is being distributed as part of a research project aimed to provide increased information about what criteria should be used in order to allocate kidney transplants, and what weight should be attached to different criteria. This is independent research, which may in time have an impact upon kidney allocation policy. It will take into account the views of key stakeholders who may have an interest in the criteria to be used to prioritize transplants. Stakeholders will include renal patients; healthcare workers routinely working with renal patients; those caring for renal patients; kidney donors and relatives of deceased donors.

In order to do this we have developed this questionnaire. The questionnaire is trying to find out your preferences for a number of different hypothetical (illustrative) health care scenarios for treatment of patients who need a renal transplantation. Please begin by familiarising yourself with the various possible alternatives by **reading the box below.** Please assume that all other factors other than those which we indicate may differ, are equal.

Background information that we can provide you with which may inform your choices is as follows:

- This questionnaire relates only to allocation from deceased (dead) donors, currently there are over 5,000 people awaiting a transplant in the UK, and this year about 1,300 will get a deceased donor transplant.

- Under the matching system currently used in the UK, it is possible to get a kidney transplant at any time after going on the list, but someone may wait for many years and occasionally someone gets a transplant after being on the list for only a few weeks.

- The main factors used to allocate kidneys are waiting time and tissue matching. This means that people with rare or unusual tissue types have reduced chance of getting a cadaveric transplant. This applies to many people from ethnic minorities.

- A difference in tissue type between the donor and recipient was the main cause of transplant rejection in the past. However, with better anti-rejection drugs, rejection is not now the main cause of transplant loss, though a kidney transplant with an excellent tissue type now has a slightly better chance of survival than one with a reasonable match (Figures from UK Transplant: 1 year survival for perfect match it is 90%; for favourable match it is 89%; and for a non-favourable match it is 86%).

Different criteria for prioritizing transplants

Description of alternatives.

A) Amount of time a person has waited.

Timescales for people receiving a transplant after being placed on a waiting list are likely to differ. The waiting time could be:

1) 1 month.
2) 2 years.
3) 10 years.

B) Tissue type matching – and likelihood of transplant success.

This affects the likelihood of a transplant proving to be successful. Below are the up to date figures from UK Transplant for the survival of all transplants in the UK. There are 6 main tissue types used in matching. A perfect tissue type match is all 6 types matching; favourable is 4-5 out of 6 matching, non-favourable is less than 4 matching. If a transplant fails the patient
will return to renal dialysis.

1) Non-favourable tissue match (86% average survival rate of the kidney 1 year after the transplant).
2) Favourable tissue match (89% average survival rate of the kidney 1 year after the transplant).
3) Perfect tissue type match (90% average survival rate of the kidney 1 year after the transplant).

C) How many dependents (either children or adults) recipients have.

Some respondents might consider that those who have dependent children, or others who are dependent either because of their age or a physical or mental handicap, ought to be prioritized for kidney transplant. So we assume that respondents might have:

1) No dependents.
2) 1 dependent.
3) 4 dependents.

D) Recipient age.

The recipient could be aged either:

1) 20 years.
2) 45 years.
3) 65 years

E) Diseases affecting life expectancy.

As well as having kidney failure, someone may have other conditions prior to kidney transplantation which affect their life expectancy. Some of the conditions which reduce life expectancy may occur in young people, and some older people may be entirely healthy apart from kidney disease. We assume these could be either:

1) None.
2) Moderate diseases (uncontrolled hypertension or obesity).
3) Severe diseases (heart attack, or stroke, or diabetes with complications).

F) Other recipient illnesses.

Someone with kidney failure may have conditions other than kidney failure which are not life-threatening but do affect their quality of life. Respondents might or might not wish to allocate kidneys according to such conditions. Examples would be:-

1) Healthy except for kidney disease.
2) Kidney disease with a condition that sometimes affects their activities, such as mild asthma.
3) Kidney disease with a condition that affects their activities on a daily basis, such as severe arthritis.

INSTRUCTIONS FOR COMPLETION OF THE QUESTIONNAIRE

- We now want you to choose between different options. Please remember that the questions are hypothetical, but we want you to assume that what the questions tells you is what is actually happening.

- Everything else about the pattern of who receives what care apart from the stated differences is identical.

- Please answer every question remembering that there are no right or wrong answers, it is finding out what your personal preferences are that matters.

- Assume there is 1 kidney that could be transplanted to either patient A or patient B.

NOW PLEASE READ DESCRIPTIONS OF OPTION A AND B AND INDICATE WHO YOU THINK SHOULD BE PRIORITISED FOR A KIDNEY TRANSPLANT – PATIENT A OR PATIENT B?:

1	Patient A	Patient B
Amount of time a person has waited for a transplant	1 month	2 years
Tissue type match – and likelihood of transplant success.	86% average 1 year chance of transplant success	89% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	No dependents	1 dependent
Recipient age	20 years	45 years
Diseases affecting life expectancy	None	Moderate: Uncontrolled hypertension or obesity
Other recipient illnesses (other than Kidney disease)	None	Mild asthma

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

2	Patient A	Patient B
Amount of time a person has waited for a transplant	2 years	10 years
Tissue type match – and likelihood of transplant success.	89% average 1 year chance of transplant success	90% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	1 dependents	4 dependents
Recipient age	45 years	65 years
Diseases affecting life expectancy	Moderate: Uncontrolled hypertension or obesity	Severe: Heart attack, or stroke, or diabetes with complications.
Other recipient illnesses (other than Kidney disease)	Mild asthma	Severe arthritis

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

3	Patient A	Patient B
Amount of time a person has waited for a transplant	10 years	1 month
Tissue type match – and likelihood of transplant success.	90% average 1 year chance of transplant success	86% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	4 dependents	No dependents
Recipient age	65 years	20 years
Diseases affecting life expectancy	Severe: Heart attack, or stroke, or diabetes with complications	None
Other recipient illnesses (other than Kidney disease)	Severe arthritis	None

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

4	Patient A	Patient B
Amount of time a person has waited for a transplant	1 month	2 years
Tissue type match – and likelihood of transplant success.	86% average 1 year chance of transplant success	89% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	1 dependents	4 dependents
Recipient age	45 years	20 years
Diseases affecting life expectancy	Moderate: Uncontrolled hypertension or obesity	Severe: Heart attack or stroke, or diabetes with complications.
Other recipient illnesses (other than Kidney disease)	Severe arthritis	None

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

5	Patient A	Patient B
Amount of time a person has waited for a transplant	2 years	10 years
Tissue type match – and likelihood of transplant success.	89% average 1 year chance of transplant success	90% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	4 dependents	No dependents
Recipient age	20 years	45 years
Diseases affecting life expectancy	Severe: Heart attack or stroke, or diabetes with complications	None
Other recipient illnesses (other than Kidney disease)	None	Mild asthma

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

6	Patient A	Patient B
Amount of time a person has waited for a transplant	10 years	1 month
Tissue type match – and likelihood of transplant success.	90% average 1 year chance of transplant success	86% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	No dependents	1 dependent
Recipient age	45 years	65 years
Diseases affecting life expectancy	None	Moderate: Uncontrolled hypertension or obesity
Other recipient illnesses (other than Kidney disease)	Mild asthma	Severe arthritis

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

7	Patient A	Patient B
Amount of time a person has waited for a transplant	1 month	2 years
Tissue type match – and likelihood of transplant success.	89% average 1 year chance of transplant success	90% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	No dependents	1 dependent
Recipient age	65 years	20 years
Diseases affecting life expectancy	Severe: Heart attack or stroke, or diabetes with complications	None
Other recipient illnesses (other than Kidney disease)	Mild asthma	Severe arthritis

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

8	Patient A	Patient B
Amount of time a person has waited for a transplant	2 years	10 years
Tissue type match – and likelihood of transplant success.	90% average 1 year chance of transplant success	86% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	1 dependent	4 dependents
Recipient age	20 years	45 years
Diseases affecting life expectancy	None	Moderate: Uncontrolled hypertension or obesity
Other recipient illnesses (other than Kidney disease)	Severe arthritis	None

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

9	Patient A	Patient B
Amount of time a person has waited for a transplant	10 years	1 month
Tissue type match – and likelihood of transplant success.	86% average 1 year chance of transplant success	89% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	4 dependents	No dependents
Recipient age	45 years	65 years
Diseases affecting life expectancy	Moderate: Uncontrolled hypertension or obesity	Severe: Heart attack or stroke, or diabetes with complications
Other recipient illnesses (other than Kidney disease)	None	Mild asthma

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

About you and your circumstances.

Gender:

(Please tick 1 box only): **Male ❑ Female ❑**

Age: ____ **years**

Do you have any children under 18 years?

(Please tick 1 box only):

Yes ❑ No ❑

If 'Yes' how many of these live in your household?

1 child ❑ 2 children ❑ 3 children ❑ More than 3 children ❑

Do you have to care for any dependent adults ***(Please tick 1 box only):***

Yes ❑ **No ❑**

If 'Yes' how many?

1 dependent adult ❑ **2 dependent adults ❑** **More than 2 dependent adults ❑**

Which of the following ethnic groups do you consider that you belong to? ***(Please tick 1 box only):***

White – British ❑

White – Irish ❑

White – Any other white background ❑
- please describe_____

Mixed – White / Black Caribbean ❑

Mixed – White / Black African ❑

Mixed – White / Asian ❑

Any other mixed background ❑
- please describe_____

Black or black British (Caribbean) ❑

Black or black British (African) ❑

Black or black British (Any other background ❑

Asian or Asian British (Indian) ❑

Asian or Asian British (Pakistani) ❑

Asian or Asian British (Bangladeshi) ❑

Asian or Asian British (Any other background) ❑

Chinese ❑

Any other ethnic group ❑
- please describe_____

Work:

Are you currently in paid employment? *(Please tick all that apply):*

Yes (Working full-time).	❑	**Yes (Working part-time)**	❑
Unemployed (Not working, but available for work).	❑	**Not working due to long-Term sickness or disability.**	❑
Retired from paid work.	❑	**Full time student**	❑
Engagement in household Duties.	❑	**Others *(please specify:)*** _____	❑

Qualifications:

What is the highest level of education you have completed *(Please tick 1 box only):*

Secondary School	❑
Vocational / Trade / College Qualification	❑
'A' level / 'AS' levels	❑
Degree level qualification(s)	❑
Other *(please describe_____).*	❑

What perspective did you adopt when answering this questionnaire?: *(Please tick 1 box only):*

Answering the questions in terms of what would be best for me	❑
Answering the questions in terms of what would be best for me and others	❑
Disregarding what is best for me and only considering what is best for others	❑

Are you a Transplanted patient? *(please tick all that apply)*:

- ❑ Currently successful.
- ❑ Now failed – If failed how long has it been since kidney failure treatment with your first dialysis / transplant:_____years_____months.

Or:

- ❑ On dialysis (not transplanted) – Since being on dialysis how long have you been on the transplant list?:_____years_____months.
- ❑ Kidney disease but not on dialysis.

Waiting list for transplantation *(if you are on a waiting list, or have been please complete the following as applicable):*

Are you on the waiting list for a kidney transplant?

Yes ❑

No ❑

If you answered 'yes' how long did you wait on the list for your most recent transplant?*(if applicable)*: _____years and:_____months.

Are you currently on kidney dialysis? If so how long have you been on kidney dialysis?*(if applicable)*:_____years and _____months.

If you are no longer on kidney dialysis but have been in the past, please indicate the total amount of time you have spent on dialysis? *(if applicable)*:_____years and _____months.

The following questions are to ask about your general health state at the moment. By placing a tick in one box in each group below, please indicate which statement best describes your own health state today.

Do not tick more than one box per question.

1. **Mobility:**
   - I have no problems in walking about ☐
   - I have some problems in walking about ☐
   - I am confined to bed ☐

2. **Self-Care:**
   - I have no problems with self-care ☐
   - I have some problems washing or dressing myself ☐
   - I am unable to wash or dress myself ☐

3. **Usual Activities** (e.g. work, study, housework, family or leisure activities)**:**
   - I have no problems with performing my usual activities ☐
   - I have some problems with performing my usual activities ☐
   - I am unable to perform my usual activities ☐

4. **Pain / Discomfort:**
   - I have no pain or discomfort ☐
   - I have moderate pain or discomfort ☐
   - I have extreme pain or discomfort ☐

5. **Anxiety / Depression:**
   - I am not anxious or depressed ☐
   - I am moderately anxious or depressed ☐
   - I am extremely anxious or depressed ☐

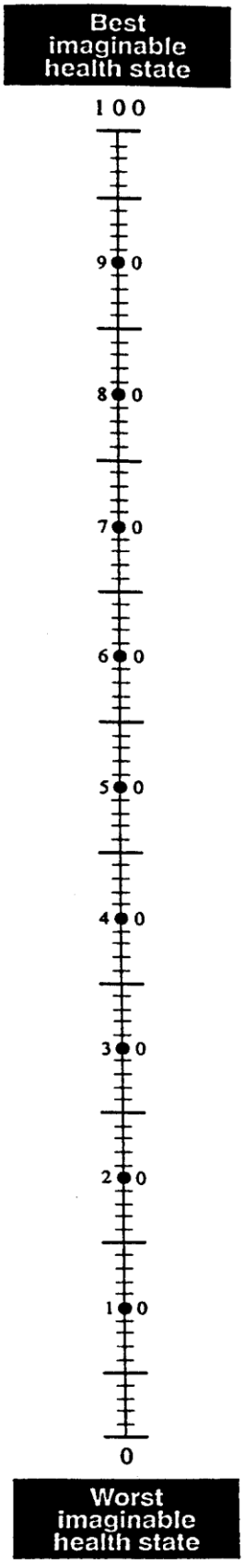
Your own health state today

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

We would like you to indicate on this scale **how good or bad is <u>your own health today</u>, in your opinion.**

Please do this by <u>drawing a line from the box below</u>, to whichever <u>point on the scale indicates how good or bad your current health state is</u> **today**.

Your own health state TODAY



REF:

Who should be prioritized for Kidney Transplants in the UK?

A SURVEY OF YOUR PREFERENCES.

Undertaken by:

Warwick Medical School,
University of Warwick,
Coventry,
UK,
CV4 7AL.

Undertaken for:

University Hospitals of Coventry
and Warwickshire NHS Trust,
Walsgrave Hospital,
Clifford Bridge Road,
Coventry,
CV2 2DX

Version 5b: Carer version: 16/05/2006

<u>What is this questionnaire about</u>?:

In this questionnaire we are trying to establish how to prioritize patients for a renal transplant. It is important for us to ensure that the systems for kidney allocation take account of the views of patients, health professionals and others involved in the transplant process. This questionnaire is being distributed as part of a research project aimed to provide increased information about what criteria should be used in order to allocate kidney transplants, and what weight should be attached to different criteria. This is independent research, which may in time have an impact upon kidney allocation policy. It will take into account the views of key stakeholders who may have an interest in the criteria to be used to prioritize transplants. Stakeholders will include renal patients; healthcare workers routinely working with renal patients; those caring for renal patients; kidney donors and relatives of deceased donors.

In order to do this we have developed this questionnaire. The questionnaire is trying to find out your preferences for a number of different hypothetical (illustrative) health care scenarios for treatment of patients who need a renal transplantation. Please begin by familiarising yourself with the various possible alternatives by **reading the box below.** Please assume that all other factors other than those which we indicate may differ, are equal.

Background information that we can provide you with which may inform your choices is as follows:

- This questionnaire relates only to allocation from deceased (dead) donors, currently there are over 5,000 people awaiting a transplant in the UK, and this year about 1,300 will get a deceased donor transplant.

- Under the matching system currently used in the UK, it is possible to get a kidney transplant at any time after going on the list, but someone may wait for many years and occasionally someone gets a transplant after being on the list for only a few weeks.

- The main factors used to allocate kidneys are waiting time and tissue matching. This means that people with rare or unusual tissue types have reduced chance of getting a cadaveric transplant. This applies to many people from ethnic minorities.

- A difference in tissue type between the donor and recipient was the main cause of transplant rejection in the past. However, with better anti-rejection drugs, rejection is not now the main cause of transplant loss, though a kidney transplant with an excellent tissue type now has a slightly better chance of survival than one with a reasonable match (Figures from UK Transplant: 1 year survival for perfect match it is 90%; for favourable match it is 89%; and for a non-favourable match it is 86%).

Different criteria for prioritizing transplants

Description of alternatives.

A) Amount of time a person has waited.

Timescales for people receiving a transplant after being placed on a waiting list are likely to differ. The waiting time could be:

1) 1 month.
2) 2 years.
3) 10 years.

B) Tissue type matching – and likelihood of transplant success.

This affects the likelihood of a transplant proving to be successful. Below are the up to date figures from UK Transplant for the survival of all transplants in the UK. There are 6 main tissue types used in matching. A perfect tissue type match is all 6 types matching; favourable is 4-5 out of 6 matching, non-favourable is less than 4 matching. If a transplant fails the patient
will return to renal dialysis.

1) Non-favourable tissue match (86% average survival rate of the kidney 1 year after the transplant).
2) Favourable tissue match (89% average survival rate of the kidney 1 year after the transplant).
3) Perfect tissue type match (90% average survival rate of the kidney 1 year after the transplant).

C) How many dependents (either children or adults) recipients have.

Some respondents might consider that those who have dependent children, or others who are dependent either because of their age or a physical or mental handicap, ought to be prioritized for kidney transplant. So we assume that respondents might have:

1) No dependents.
2) 1 dependent.
3) 4 dependents.

D) Recipient age.

The recipient could be aged either:

1) 20 years.
2) 45 years.
3) 65 years

E) Diseases affecting life expectancy.

As well as having kidney failure, someone may have other conditions prior to kidney transplantation which affect their life expectancy. Some of the conditions which reduce life expectancy may occur in young people, and some older people may be entirely

healthy apart from kidney disease. We assume these could be either:

1) None.
2) Moderate diseases (uncontrolled hypertension or obesity).
3) Severe diseases (heart attack, or stroke, or diabetes with complications).

F) Other recipient illnesses.

Someone with kidney failure may have conditions other than kidney failure which are not life-threatening but do affect their quality of life. Respondents might or might not wish to allocate kidneys according to such conditions. Examples would be:-

1) Healthy except for kidney disease.
2) Kidney disease with a condition that sometimes affects their activities, such as mild asthma.
3) Kidney disease with a condition that affects their activities on a daily basis, such as severe arthritis.

INSTRUCTIONS FOR COMPLETION OF THE QUESTIONNAIRE

- We now want you to choose between different options. Please remember that the questions are hypothetical, but we want you to assume that what the questions tells you is what is actually happening.

- Everything else about the pattern of who receives what care apart from the stated differences is identical.

- Please answer every question remembering that there are no right or wrong answers, it is finding out what your personal preferences are that matters.

- Assume there is 1 kidney that could be transplanted to either patient A or patient B.

NOW PLEASE READ DESCRIPTIONS OF OPTION A AND B AND INDICATE WHO YOU THINK SHOULD BE PRIORITISED FOR A KIDNEY TRANSPLANT – PATIENT A OR PATIENT B?:

1	Patient A	Patient B
Amount of time a person has waited for a transplant	1 month	2 years
Tissue type match – and likelihood of transplant success.	90% average 1 year chance of transplant success	86% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	4 dependents	No dependents
Recipient age	20 years	45 years
Diseases affecting life expectancy	Moderate: Uncontrolled hypertension or obesity	Severe: Heart attack, or stroke, or diabetes with complications.
Other recipient illnesses (other than Kidney disease)	Mild asthma	Severe arthritis

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

2	Patient A	Patient B
Amount of time a person has waited for a transplant	2 years	10 years
Tissue type match – and likelihood of transplant success.	86% average 1 year chance of transplant success	89% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	No dependents	1 dependent
Recipient age	45 years	65 years
Diseases affecting life expectancy	Severe: Heart attack, or stroke, or diabetes with complications	None
Other recipient illnesses (other than Kidney disease)	Severe arthritis	None

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

3	Patient A	Patient B
Amount of time a person has waited for a transplant	10 years	1 month
Tissue type match – and likelihood of transplant success.	89% average 1 year chance of transplant success	90% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	1 dependent	4 dependents
Recipient age	65 years	20 years
Diseases affecting life expectancy	None	Moderate: Uncontrolled hypertension or obesity
Other recipient illnesses (other than Kidney disease)	None	Mild asthma

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

4	Patient A	Patient B
Amount of time a person has waited for a transplant	1 month	2 years
Tissue type match – and likelihood of transplant success.	89% average 1 year chance of transplant success	90% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	4 dependents	No dependents
Recipient age	45 years	65 years
Diseases affecting life expectancy	None	Moderate: Uncontrolled hypertension or obesity
Other recipient illnesses (other than Kidney disease)	Severe arthritis	None

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

5	Patient A	Patient B
Amount of time a person has waited for a transplant	2 years	10 years
Tissue type match – and likelihood of transplant success.	90% average 1 year chance of transplant success	86% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	No dependents	1 dependent
Recipient age	65 years	20 years
Diseases affecting life expectancy	Moderate: Uncontrolled hypertension or obesity	Severe: Heart attack, or stroke, or diabetes with complications.
Other recipient illnesses (other than Kidney disease)	None	Mild asthma

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

6	Patient A	Patient B
Amount of time a person has waited for a transplant	10 years	1 month
Tissue type match – and likelihood of transplant success.	86% average 1 year chance of transplant success	86% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	1 dependent	4 dependents
Recipient age	20 years	45 years
Diseases affecting life expectancy	Severe: Heart attack, or stroke, or diabetes with complications.	None
Other recipient illnesses (other than Kidney disease)	Mild asthma	Severe arthritis

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

7	Patient A	Patient B
Amount of time a person has waited for a transplant	1 month	2 years
Tissue type match – and likelihood of transplant success.	90% average 1 year chance of transplant success	86% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	1 dependent	4 dependents
Recipient age	45 years	65 years
Diseases affecting life expectancy	Severe: Heart attack or stroke, or diabetes with complications	None
Other recipient illnesses (other than Kidney disease)	None	Mild asthma

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

8	Patient A	Patient B
Amount of time a person has waited for a transplant	2 years	10 years
Tissue type match – and likelihood of transplant success.	86% average 1 year chance of transplant success	89% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	4 dependents	No dependents
Recipient age	65 years	20 years
Diseases affecting life expectancy	None	Moderate: Uncontrolled hypertension or obesity
Other recipient illnesses (other than Kidney disease)	Mild asthma	Severe arthritis

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

9	Patient A	Patient B
Amount of time a person has waited for a transplant	10 years	1 month
Tissue type match – and likelihood of transplant success.	89% average 1 year chance of transplant success	90% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	No dependents	1 dependent
Recipient age	20 years	45 years
Diseases affecting life expectancy	Moderate: Uncontrolled hypertension or obesity	Severe: Heart attack or stroke, or diabetes with complications
Other recipient illnesses (other than Kidney disease)	Severe arthritis.	None

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

About you and your circumstances.

Gender:

(Please tick 1 box only): **Male ❑ Female ❑**

Age: ____ **years**

Do you have any children under 18 years?

(Please tick 1 box only):

Yes ❑ No ❑

If 'Yes' how many of these live in your household?

1 child ❑ 2 children ❑ 3 children ❑ More than 3 children ❑

Do you have to care for any dependent adults ***(Please tick 1 box only):***

Yes ❑ **No ❑**

If 'Yes' how many?

1 dependent adult ❑ **2 dependent adults ❑** **More than 2 dependent adults ❑**

Which of the following ethnic groups do you consider that you belong to? ***(Please tick 1 box only):***

White – British ❑

White – Irish ❑

White – Any other white background ❑
- please describe_____

Mixed – White / Black Caribbean ❑

Mixed – White / Black African ❑

Mixed – White / Asian ❑

Any other mixed background ❑
- please describe_____

Black or black British (Caribbean) ❑

Black or black British (African) ❑

Black or black British (Any other background ❑

Asian or Asian British (Indian) ❑

Asian or Asian British (Pakistani) ❑

Asian or Asian British (Bangladeshi) ❑

Asian or Asian British (Any other background) ❑

Chinese ❑

Any other ethnic group ❑
- please describe_____

<u>Work</u>:

Are you currently in paid employment? *(Please tick all that apply):*

Yes (Working full-time).	❑	**Yes (Working part-time)**	❑
Unemployed (Not working, but available for work).	❑	**Not working due to long-Term sickness or disability.**	❑
Retired from paid work.	❑	**Full time student**	❑
Engagement in household Duties.	❑	**Others *(please specify:)*** _____	❑

<u>Qualifications</u>:

What is the highest level of education you have completed *(Please tick 1 box only):*

Secondary School	❑
Vocational / Trade / College Qualification	❑
'A' level / 'AS' levels	❑
Degree level qualification(s)	❑
Other *(please describe*_____*).*	❑

<u>What perspective did you adopt when answering this questionnaire?:</u> *(Please tick 1 box only):*

Answering the questions in terms of what would be best for me	❑
Answering the questions in terms of what would be best for me and others	❑
Disregarding what is best for me and only considering what is best for others	❑

Which of the following categories does the person you are caring for fall into:

The person I am caring for is a transplanted patient *(please tick all that apply)*:

- ❑ Currently successful.
- ❑ Now failed – If failed how long has it been since kidney failure treatment with your first dialysis / transplant:_____years_____months.

Or:

- ❑ On dialysis – Since being on dialysis how long have they been on the transplant list?:_____years_____months.
- ❑ Kidney disease but not on dialysis.

Is the person you are caring for on the waiting list for transplantation?

Yes ❑

No ❑

If you answered 'yes' how long did she / he wait on the list for their most recent transplant?*(if applicable)*: _____years and:_____months.

Are they currently on kidney dialysis? If so how long have they been on kidney dialysis?*(if applicable)*:_____years and _____months.

If they are no longer on kidney dialysis but have been in the past, please indicate the total amount of time they have spent on dialysis? *(if applicable)*:_____years and _____months.

REF:

Who should be prioritized for Kidney Transplants in the UK?

A SURVEY OF YOUR PREFERENCES.

Undertaken by:

Warwick Medical School,
University of Warwick,
Coventry,
UK,
CV4 7AL.

Undertaken for:

University Hospitals of Coventry
and Warwickshire NHS Trust,
Walsgrave Hospital,
Clifford Bridge Road,
Coventry,
CV2 2DX

Version 6a: Healthcare professional version: 16/05/2006

What is this questionnaire about?:

In this questionnaire we are trying to establish how to prioritize patients for a renal transplant. It is important for us to ensure that the systems for kidney allocation take account of the views of patients, health professionals and others involved in the transplant process. This questionnaire is being distributed as part of a research project aimed to provide increased information about what criteria should be used in order to allocate kidney transplants, and what weight should be attached to different criteria. This is independent research, which may in time have an impact upon kidney allocation policy. It will take into account the views of key stakeholders who may have an interest in the criteria to be used to prioritize transplants. Stakeholders will include renal patients; healthcare workers routinely working with renal patients; those caring for renal patients; kidney donors and relatives of deceased donors.

In order to do this we have developed this questionnaire. The questionnaire is trying to find out your preferences for a number of different hypothetical (illustrative) health care scenarios for treatment of patients who need a renal transplantation. Please begin by familiarising yourself with the various possible alternatives by **reading the box below.** Please assume that all other factors other than those which we indicate may differ, are equal.

Background information that we can provide you with which may inform your choices is as follows:

- This questionnaire relates only to allocation from deceased (dead) donors, currently there are over 5,000 people awaiting a transplant in the UK, and this year about 1,300 will get a deceased donor transplant.

- Under the matching system currently used in the UK, it is possible to get a kidney transplant at any time after going on the list, but someone may wait for many years and occasionally someone gets a transplant after being on the list for only a few weeks.

- The main factors used to allocate kidneys are waiting time and tissue matching. This means that people with rare or unusual tissue types have reduced chance of getting a cadaveric transplant. This applies to many people from ethnic minorities.

- A difference in tissue type between the donor and recipient was the main cause of transplant rejection in the past. However, with better anti-rejection drugs, rejection is not now the main cause of transplant loss, though a kidney transplant with an excellent tissue type now has a slightly better chance of survival than one with a reasonable match (Figures from UK Transplant: 1 year survival for perfect match it is 90%; for favourable match it is 89%; and for a non-favourable match it is 86%).

Different criteria for prioritizing transplants

Description of alternatives.

A) Amount of time a person has waited.

Timescales for people receiving a transplant after being placed on a waiting list are likely to differ. The waiting time could be:

1) 1 month.
2) 2 years.
3) 10 years.

B) Tissue type matching – and likelihood of transplant success.

This affects the likelihood of a transplant proving to be successful. Below are the up to date figures from UK Transplant for the survival of all transplants in the UK. There are 6 main tissue types used in matching. A perfect tissue type match is all 6 types matching; favourable is 4-5 out of 6 matching, non-favourable is less than 4 matching. If a transplant fails the patient
will return to renal dialysis.

1) Non-favourable tissue match (86% average survival rate of the kidney 1 year after the transplant).
2) Favourable tissue match (89% average survival rate of the kidney 1 year after the transplant).
3) Perfect tissue type match (90% average survival rate of the kidney 1 year after the transplant).

C) How many dependents (either children or adults) recipients have.

Some respondents might consider that those who have dependent children, or others who are dependent either because of their age or a physical or mental handicap, ought to be prioritized for kidney transplant. So we assume that respondents might have:

1) No dependents.
2) 1 dependent.
3) 4 dependents.

D) Recipient age.

The recipient could be aged either:

1) 20 years.
2) 45 years.
3) 65 years

E) Diseases affecting life expectancy.

As well as having kidney failure, someone may have other conditions prior to kidney transplantation which affect their life expectancy. Some of the conditions which reduce life expectancy may occur in young people, and some older people may be entirely

healthy apart from kidney disease. We assume these could be either:

1) None.
2) Moderate diseases (uncontrolled hypertension or obesity).
3) Severe diseases (heart attack, or stroke, or diabetes with complications).

F) Other recipient illnesses.

Someone with kidney failure may have conditions other than kidney failure which are not life-threatening but do affect their quality of life. Respondents might or might not wish to allocate kidneys according to such conditions. Examples would be:-

1) Healthy except for kidney disease.
2) Kidney disease with a condition that sometimes affects their activities, such as mild asthma.
3) Kidney disease with a condition that affects their activities on a daily basis, such as severe arthritis.

INSTRUCTIONS FOR COMPLETION OF THE QUESTIONNAIRE

- We now want you to choose between different options. Please remember that the questions are hypothetical, but we want you to assume that what the questions tells you is what is actually happening.

- Everything else about the pattern of who receives what care apart from the stated differences is identical.

- Please answer every question remembering that there are no right or wrong answers, it is finding out what your personal preferences are that matters.

- Assume there is 1 kidney that could be transplanted to either patient A or patient B.

NOW PLEASE READ DESCRIPTIONS OF OPTION A AND B AND INDICATE WHO YOU THINK SHOULD BE PRIORITISED FOR A KIDNEY TRANSPLANT – PATIENT A OR PATIENT B?:

1	Patient A	Patient B
Amount of time a person has waited for a transplant	1 month	2 years
Tissue type match – and likelihood of transplant success.	86% average 1 year chance of transplant success	89% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	No dependents	1 dependent
Recipient age	20 years	45 years
Diseases affecting life expectancy	None	Moderate: Uncontrolled hypertension or obesity
Other recipient illnesses (other than Kidney disease)	None	Mild asthma

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

2	Patient A	Patient B
Amount of time a person has waited for a transplant	2 years	10 years
Tissue type match – and likelihood of transplant success.	89% average 1 year chance of transplant success	90% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	1 dependents	4 dependents
Recipient age	45 years	65 years
Diseases affecting life expectancy	Moderate: Uncontrolled hypertension or obesity	Severe: Heart attack, or stroke, or diabetes with complications.
Other recipient illnesses (other than Kidney disease)	Mild asthma	Severe arthritis

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

3	Patient A	Patient B
Amount of time a person has waited for a transplant	10 years	1 month
Tissue type match – and likelihood of transplant success.	90% average 1 year chance of transplant success	86% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	4 dependents	No dependents
Recipient age	65 years	20 years
Diseases affecting life expectancy	Severe: Heart attack, or stroke, or diabetes with complications	None
Other recipient illnesses (other than Kidney disease)	Severe arthritis	None

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

4	Patient A	Patient B
Amount of time a person has waited for a transplant	1 month	2 years
Tissue type match – and likelihood of transplant success.	86% average 1 year chance of transplant success	89% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	1 dependents	4 dependents
Recipient age	45 years	20 years
Diseases affecting life expectancy	Moderate: Uncontrolled hypertension or obesity	Severe: Heart attack or stroke, or diabetes with complications.
Other recipient illnesses (other than Kidney disease)	Severe arthritis	None

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

5	Patient A	Patient B
Amount of time a person has waited for a transplant	2 years	10 years
Tissue type match – and likelihood of transplant success.	89% average 1 year chance of transplant success	90% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	4 dependents	No dependents
Recipient age	20 years	45 years
Diseases affecting life expectancy	Severe: Heart attack or stroke, or diabetes with complications	None
Other recipient illnesses (other than Kidney disease)	None	Mild asthma

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

6	Patient A	Patient B
Amount of time a person has waited for a transplant	10 years	1 month
Tissue type match – and likelihood of transplant success.	90% average 1 year chance of transplant success	86% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	No dependents	1 dependent
Recipient age	45 years	65 years
Diseases affecting life expectancy	None	Moderate: Uncontrolled hypertension or obesity
Other recipient illnesses (other than Kidney disease)	Mild asthma	Severe arthritis

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

7	Patient A	Patient B
Amount of time a person has waited for a transplant	1 month	2 years
Tissue type match – and likelihood of transplant success.	89% average 1 year chance of transplant success	90% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	No dependents	1 dependent
Recipient age	65 years	20 years
Diseases affecting life expectancy	Severe: Heart attack or stroke, or diabetes with complications	None
Other recipient illnesses (other than Kidney disease)	Mild asthma	Severe arthritis

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

8	Patient A	Patient B
Amount of time a person has waited for a transplant	2 years	10 years
Tissue type match – and likelihood of transplant success.	90% average 1 year chance of transplant success	86% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	1 dependent	4 dependents
Recipient age	20 years	45 years
Diseases affecting life expectancy	None	Moderate: Uncontrolled hypertension or obesity
Other recipient illnesses (other than Kidney disease)	Severe arthritis	None

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

9	Patient A	Patient B
Amount of time a person has waited for a transplant	10 years	1 month
Tissue type match – and likelihood of transplant success.	86% average 1 year chance of transplant success	89% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	4 dependents	No dependents
Recipient age	45 years	65 years
Diseases affecting life expectancy	Moderate: Uncontrolled hypertension or obesity	Severe: Heart attack or stroke, or diabetes with complications
Other recipient illnesses (other than Kidney disease)	None	Mild asthma

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

About you and your circumstances.

Gender:

(Please tick 1 box only): **Male ❑ Female ❑**

Age: ____ **years**

Do you have any children under 18 years?

(Please tick 1 box only):

Yes ❑ No ❑

If 'Yes' how many of these live in your household?

1 child ❑ 2 children ❑ 3 children ❑ More than 3 children ❑

Do you have to care for any dependent adults ***(Please tick 1 box only):***

Yes ❑ **No ❑**

If 'Yes' how many?

1 dependent adult ❑ **2 dependent adults ❑** **More than 2 dependent adults ❑**

Which of the following ethnic groups do you consider that you belong to? ***(Please tick 1 box only):***

White – British ❑

White – Irish ❑

White – Any other white background ❑
- please describe_____

Mixed – White / Black Caribbean ❑

Mixed – White / Black African ❑

Mixed – White / Asian ❑

Any other mixed background ❑
- please describe_____

Black or black British (Caribbean) ❑

Black or black British (African) ❑

Black or black British (Any other background ❑

Asian or Asian British (Indian) ❑

Asian or Asian British (Pakistani) ❑

Asian or Asian British (Bangladeshi) ❑

Asian or Asian British (Any other background) ❑

Chinese ❑

Any other ethnic group ❑
- please describe_____

<u>Work</u>:

Are you currently in paid employment? *(Please tick all that apply):*

Yes (Working full-time).	❑	**Yes (Working part-time)**	❑
Unemployed (Not working, but available for work).	❑	**Not working due to long-Term sickness or disability.**	❑
Retired from paid work.	❑	**Full time student**	❑
Engagement in household Duties.	❑	**Others *(please specify:)*** _____	❑

<u>Qualifications</u>:

What is the highest level of education you have completed *(Please tick 1 box only):*

Secondary School	❑
Vocational / Trade / College Qualification	❑
'A' level / 'AS' levels	❑
Degree level qualification(s)	❑
Other *(please describe*_____*).*	❑

<u>What perspective did you adopt when answering this questionnaire?:</u> *(Please tick 1 box only):*

Answering the questions in terms of what would be best for me	❑
Answering the questions in terms of what would be best for me and others	❑
Disregarding what is best for me and only considering what is best for others	❑

Which of the following categories do you fall into:

- ❑ Surgeon.
- ❑ Renal Physician.
- ❑ Transplant co-ordinator.
- ❑ Nurse.
- ❑ Pharmacist.
- ❑ **Other *(please describe)*_____*.***

REF:

Who should be prioritized for Kidney Transplants in the UK?

A SURVEY OF YOUR PREFERENCES.

Undertaken by:

Warwick Medical School,
University of Warwick,
Coventry,
UK,
CV4 7AL.

Undertaken for:

University Hospitals of Coventry
and Warwickshire NHS Trust,
Walsgrave Hospital,
Clifford Bridge Road,
Coventry,
CV2 2DX

Version 7b: Donor / relative of deceased donor version: 04/04/2006

<u>What is this questionnaire about</u>?:

In this questionnaire we are trying to establish how to prioritize patients for a renal transplant. It is important for us to ensure that the systems for kidney allocation take account of the views of patients, health professionals and others involved in the transplant process. This questionnaire is being distributed as part of a research project aimed to provide increased information about what criteria should be used in order to allocate kidney transplants, and what weight should be attached to different criteria. This is independent research, which may in time have an impact upon kidney allocation policy. It will take into account the views of key stakeholders who may have an interest in the criteria to be used to prioritize transplants. Stakeholders will include renal patients; healthcare workers routinely working with renal patients; those caring for renal patients; kidney donors and relatives of deceased donors.

In order to do this we have developed this questionnaire. The questionnaire is trying to find out your preferences for a number of different hypothetical (illustrative) health care scenarios for treatment of patients who need a renal transplantation. Please begin by familiarising yourself with the various possible alternatives by **reading the box below.** Please assume that all other factors other than those which we indicate may differ, are equal.

Background information that we can provide you with which may inform your choices is as follows:

- This questionnaire relates only to allocation from deceased (dead) donors, currently there are over 5,000 people awaiting a transplant in the UK, and this year about 1,300 will get a deceased donor transplant.

- Under the matching system currently used in the UK, it is possible to get a kidney transplant at any time after going on the list, but someone may wait for many years and occasionally someone gets a transplant after being on the list for only a few weeks.

- The main factors used to allocate kidneys are waiting time and tissue matching. This means that people with rare or unusual tissue types have reduced chance of getting a cadaveric transplant. This applies to many people from ethnic minorities.

- A difference in tissue type between the donor and recipient was the main cause of transplant rejection in the past. However, with better anti-rejection drugs, rejection is not now the main cause of transplant loss, though a kidney transplant with an excellent tissue type now has a slightly better chance of survival than one with a reasonable match (Figures from UK Transplant: 1 year survival for perfect match it is 90%; for favourable match it is 89%; and for a non-favourable match it is 86%).

Different criteria for prioritizing transplants

Description of alternatives.

A) Amount of time a person has waited.

Timescales for people receiving a transplant after being placed on a waiting list are likely to differ. The waiting time could be:

1) 1 month.
2) 2 years.
3) 10 years.

B) Tissue type matching – and likelihood of transplant success.

This affects the likelihood of a transplant proving to be successful. Below are the up to date figures from UK Transplant for the survival of all transplants in the UK. There are 6 main tissue types used in matching. A perfect tissue type match is all 6 types matching; favourable is 4-5 out of 6 matching, non-favourable is less than 4 matching. If a transplant fails the patient
will return to renal dialysis.

1) Non-favourable tissue match (86% average survival rate of the kidney 1 year after the transplant).
2) Favourable tissue match (89% average survival rate of the kidney 1 year after the transplant).
3) Perfect tissue type match (90% average survival rate of the kidney 1 year after the transplant).

C) How many dependents (either children or adults) recipients have.

Some respondents might consider that those who have dependent children, or others who are dependent either because of their age or a physical or mental handicap, ought to be prioritized for kidney transplant. So we assume that respondents might have:

1) No dependents.
2) 1 dependent.
3) 4 dependents.

D) Recipient age.

The recipient could be aged either:

1) 20 years.
2) 45 years.
3) 65 years

E) Diseases affecting life expectancy.

As well as having kidney failure, someone may have other conditions prior to kidney transplantation which affect their life expectancy. Some of the conditions which reduce life expectancy may occur in young people, and some older people may be entirely

healthy apart from kidney disease. We assume these could be either:

1) None.
2) Moderate diseases (uncontrolled hypertension or obesity).
3) Severe diseases (heart attack, or stroke, or diabetes with complications).

F) Other recipient illnesses.

Someone with kidney failure may have conditions other than kidney failure which are not life-threatening but do affect their quality of life. Respondents might or might not wish to allocate kidneys according to such conditions. Examples would be:-

1) Healthy except for kidney disease.
2) Kidney disease with a condition that sometimes affects their activities, such as mild asthma.
3) Kidney disease with a condition that affects their activities on a daily basis, such as severe arthritis.

INSTRUCTIONS FOR COMPLETION OF THE QUESTIONNAIRE

- We now want you to choose between different options. Please remember that the questions are hypothetical, but we want you to assume that what the questions tells you is what is actually happening.

- Everything else about the pattern of who receives what care apart from the stated differences is identical.

- Please answer every question remembering that there are no right or wrong answers, it is finding out what your personal preferences are that matters.

- Assume there is 1 kidney that could be transplanted to either patient A or patient B.

NOW PLEASE READ DESCRIPTIONS OF OPTION A AND B AND INDICATE WHO YOU THINK SHOULD BE PRIORITISED FOR A KIDNEY TRANSPLANT – PATIENT A OR PATIENT B?:

1	Patient A	Patient B
Amount of time a person has waited for a transplant	1 month	2 years
Tissue type match – and likelihood of transplant success.	90% average 1 year chance of transplant success	86% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	4 dependents	No dependents
Recipient age	20 years	45 years
Diseases affecting life expectancy	Moderate: Uncontrolled hypertension or obesity	Severe: Heart attack, or stroke, or diabetes with complications.
Other recipient illnesses (other than Kidney disease)	Mild asthma	Severe arthritis

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

2	Patient A	Patient B
Amount of time a person has waited for a transplant	2 years	10 years
Tissue type match – and likelihood of transplant success.	86% average 1 year chance of transplant success	89% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	No dependents	1 dependent
Recipient age	45 years	65 years
Diseases affecting life expectancy	Severe: Heart attack, or stroke, or diabetes with complications	None
Other recipient illnesses (other than Kidney disease)	Severe arthritis	None

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

3	Patient A	Patient B
Amount of time a person has waited for a transplant	10 years	1 month
Tissue type match – and likelihood of transplant success.	89% average 1 year chance of transplant success	90% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	1 dependent	4 dependents
Recipient age	65 years	20 years
Diseases affecting life expectancy	None	Moderate: Uncontrolled hypertension or obesity
Other recipient illnesses (other than Kidney disease)	None	Mild asthma

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

4	Patient A	Patient B
Amount of time a person has waited for a transplant	1 month	2 years
Tissue type match – and likelihood of transplant success.	89% average 1 year chance of transplant success	90% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	4 dependents	No dependents
Recipient age	45 years	65 years
Diseases affecting life expectancy	None	Moderate: Uncontrolled hypertension or obesity
Other recipient illnesses (other than Kidney disease)	Severe arthritis	None

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

5	Patient A	Patient B
Amount of time a person has waited for a transplant	2 years	10 years
Tissue type match – and likelihood of transplant success.	90% average 1 year chance of transplant success	86% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	No dependents	1 dependent
Recipient age	65 years	20 years
Diseases affecting life expectancy	Moderate: Uncontrolled hypertension or obesity	Severe: Heart attack, or stroke, or diabetes with complications.
Other recipient illnesses (other than Kidney disease)	None	Mild asthma

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

6	Patient A	Patient B
Amount of time a person has waited for a transplant	10 years	1 month
Tissue type match – and likelihood of transplant success.	86% average 1 year chance of transplant success	86% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	1 dependent	4 dependents
Recipient age	20 years	45 years
Diseases affecting life expectancy	Severe: Heart attack, or stroke, or diabetes with complications.	None
Other recipient illnesses (other than Kidney disease)	Mild asthma	Severe arthritis

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

7	Patient A	Patient B
Amount of time a person has waited for a transplant	1 month	2 years
Tissue type match – and likelihood of transplant success.	90% average 1 year chance of transplant success	86% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	1 dependent	4 dependents
Recipient age	45 years	65 years
Diseases affecting life expectancy	Severe: Heart attack or stroke, or diabetes with complications	None
Other recipient illnesses (other than Kidney disease)	None	Mild asthma

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

8	Patient A	Patient B
Amount of time a person has waited for a transplant	2 years	10 years
Tissue type match – and likelihood of transplant success.	86% average 1 year chance of transplant success	89% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	4 dependents	No dependents
Recipient age	65 years	20 years
Diseases affecting life expectancy	None	Moderate: Uncontrolled hypertension or obesity
Other recipient illnesses (other than Kidney disease)	Mild asthma	Severe arthritis

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

9	Patient A	Patient B
Amount of time a person has waited for a transplant	10 years	1 month
Tissue type match – and likelihood of transplant success.	89% average 1 year chance of transplant success	90% average 1 year chance of transplant success
How many dependents (children or adults) recipients have.	No dependents	1 dependent
Recipient age	20 years	45 years
Diseases affecting life expectancy	Moderate: Uncontrolled hypertension or obesity	Severe: Heart attack or stroke, or diabetes with complications
Other recipient illnesses (other than Kidney disease)	Severe arthritis.	None

Which patient would you choose? (tick 1 box only) Patient A ❑ Patient B ❑

About you and your circumstances.

Gender:

(Please tick 1 box only): **Male ❑ Female ❑**

Age: ____ **years**

Do you have any children under 18 years?

(Please tick 1 box only):

Yes ❑ No ❑

If 'Yes' how many of these live in your household?

1 child ❑ 2 children ❑ 3 children ❑ More than 3 children ❑

Do you have to care for any dependent adults ***(Please tick 1 box only):***

Yes ❑ **No ❑**

If 'Yes' how many?

1 dependent adult ❑ **2 dependent adults ❑** **More than 2 dependent adults ❑**

Which of the following ethnic groups do you consider that you belong to? ***(Please tick 1 box only):***

White – British ❑

White – Irish ❑

White – Any other white background ❑
- please describe_____

Mixed – White / Black Caribbean ❑

Mixed – White / Black African ❑

Mixed – White / Asian ❑

Any other mixed background ❑
- please describe_____

Black or black British (Caribbean) ❑

Black or black British (African) ❑

Black or black British (Any other background ❑

Asian or Asian British (Indian) ❑

Asian or Asian British (Pakistani) ❑

Asian or Asian British (Bangladeshi) ❑

Asian or Asian British (Any other background) ❑

Chinese ❑

Any other ethnic group ❑
- please describe_____

<u>Work</u>:

Are you currently in paid employment? *(Please tick all that apply):*

Yes (Working full-time).	❑	**Yes (Working part-time)**	❑
Unemployed (Not working, but available for work).	❑	**Not working due to long-Term sickness or disability.**	❑
Retired from paid work.	❑	**Full time student**	❑
Engagement in household Duties.	❑	**Others *(please specify:)*** _____	❑

<u>Qualifications</u>:

What is the highest level of education you have completed *(Please tick 1 box only):*

Secondary School	❑
Vocational / Trade / College Qualification	❑
'A' level / 'AS' levels	❑
Degree level qualification(s)	❑
Other *(please describe*_____*).*	❑

<u>What perspective did you adopt when answering this questionnaire?:</u> *(Please tick 1 box only):*

Answering the questions in terms of what would be best for me	❑
Answering the questions in terms of what would be best for me and others	❑
Disregarding what is best for me and only considering what is best for others	❑

Are you a donor or relative of a deceased donor?:

- ❑ A living donor.
- ❑ A relative of a deceased donor.